"十二五"普通高等教育本科国家级规划教材

卫 生 部 "十 二 五" 规 划 教 材

全国高等医药教材建设研究会"十二五"规划教材

全国高等学校教材

供基础、临床、预防、口腔医学类专业用

临床流行病学与循证医学

Clinical Epidemiology and Evidence Based Medicine

第4版

主　审　王家良　王滨有

主　编　刘续宝　王素萍

副主编　孙业桓　时景璞

人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

网络增值服务

ONLINE SERVICES

“十二五”普通高等教育本科国家级规划教材

卫生部“十二五”规划教材
全国高等医药教材建设研究会“十二五”规划教材
全国高等学校教材

供基础、临床、预防、口腔医学类专业用

临床流行病学与循证医学

Clinical Epidemiology and Evidence Based Medicine

第4版

主　审　王家良　王滨有

主　编　刘续宝　王素萍

副主编　孙业桓　时景璞

编　者（按姓氏笔画排序）

王小钦（复旦大学）
王素萍（山西医科大学）
方　芳（四川大学）
吕　明（山东大学）
刘晓清（北京协和医学院）
刘续宝（四川大学）
闫永平（第四军医大学）
孙业桓（安徽医科大学）
李　革（重庆医科大学）
李　静（四川大学）
时景璞（中国医科大学）
陈世耀（复旦大学）
赵亚双（哈尔滨医科大学）
洪明晃（中山大学）
黄悦勤（北京大学）
康德英（四川大学）
廖晓阳（四川大学）

秘　书　康德英　洪　旗（四川大学）

人民卫生出版社

图书在版编目(CIP)数据

临床流行病学与循证医学/刘续宝等主编. —4版. —北京：人民卫生出版社，2013

ISBN 978-7-117-17062-8

Ⅰ. ①临… Ⅱ. ①刘… Ⅲ. ①临床流行病学—医学院校—教材②临床医学—医学院校—教材 Ⅳ. ①R181.3②R4

中国版本图书馆CIP数据核字(2013)第051011号

人卫社官网	www.pmph.com	出版物查询，在线购书
人卫医学网	www.ipmph.com	医学考试辅导，医学数据库服务，医学教育资源，大众健康资讯

临床流行病学与循证医学

第4版

主　　编：刘续宝　王素萍
出版发行：人民卫生出版社（中继线 010-59780011）
地　　址：北京市朝阳区潘家园南里19号
邮　　编：100021
E - mail：pmph @ pmph.com
购书热线：010-59787592　010-59787584　010-65264830
印　　刷：北京市卫顺印刷厂
经　　销：新华书店
开　　本：850×1168　1/16　印张：16　插页：9
字　　数：440千字
版　　次：2002年8月第1版　2013年3月第4版
　　　　　2013年3月第4版第1次印刷（总第10次印刷）
标准书号：ISBN 978-7-117-17062-8/R · 17063
定　　价：38.00元
打击盗版举报电话：010-59787491　E-mail：WQ @ pmph.com
（凡属印装质量问题请与本社市场营销中心联系退换）

全国高等学校五年制本科临床医学专业
第八轮 教材修订说明

全国高等学校五年制本科临床医学专业卫生部规划教材自 1978 年第一轮出版至今已有 35 年的历史。几十年来，在教育部、卫生部的领导和支持下，以裘法祖、吴阶平、吴孟超、陈灏珠等院士为代表的我国几代德高望重、有丰富的临床和教学经验、有高度责任感和敬业精神的国内外著名院士、专家、医学家、教育家参与了本套教材的创建和每一轮教材的修订工作，使我国的五年制本科临床医学教材从无到有，从少到多，从多到精，不断丰富、完善与创新，形成了课程门类齐全、学科系统优化、内容衔接合理、结构体系科学的由规划教材、配套教材、配套光盘、数字出版、网络增值服务组成的立体化教材格局。这套教材为我国千百万医学生的培养和成才提供了根本保障，为我国培养了一代又一代高水平、高素质的合格医学人才，为推动我国医疗卫生事业的改革和发展做出了历史性巨大贡献，并通过教材的创新建设和高质量发展，推动了我国高等医学本科教育的改革和发展，促进了我国医药学相关学科或领域的教材建设和教育发展，走出了一条适合中国医药学教育和卫生事业发展实际的具有中国特色医药学教材建设和发展的道路，创建了中国特色医药学教育教材建设模式。老一辈医学教育家和科学家们亲切地称这套教材是中国医学教育的“干细胞”教材。

本套第八轮教材修订启动之时，正是全球医学教育百年反思之际，更是我国医疗卫生体制改革和医学教育改革全方位深入推进之时，教育部、卫生部共同召开了全国医学教育改革工作会议，启动了“5+3”为主体的临床医学教育综合改革，形成了以医改推动教改，教改服务医改的历史发展格局。人民卫生出版社和全国高等医药教材建设研究会紧紧抓住医学教育综合改革的历史发展机遇期，以全国高等学校五年制本科临床医学专业第八轮规划教材全面启动为契机，以规划教材创新建设，全面推进国家级规划教材建设工作，服务于医改和教改。

第八轮教材的修订原则是积极贯彻落实教育部、卫生部关于实施临床医学教育综合改革的意见，努力优化人才培养结构，坚持以需求为导向，构建发展以“5+3”模式为主体的临床医学人才培养体系；改革课程体系、教学内容、教学方法和评价考核办法；将医德教育贯穿于医学教育的全过程，强化临床实践教学，采取多种措施，切实落实好“早临床、多临床、反复临床”的要求，提高医学生的临床实践能力。

在全国医学教育综合改革精神鼓舞下和老一辈医学家奉献精神的感召下，全国一大批临床教学、科研、医疗第一线的中青年专家、学者、教授继承和发扬了老一辈的优秀传统，以严谨治学的科学态度和无私奉献的敬业精神，积极参与第八轮教材的修订和建设工作，紧密结合五年制临床医学专业培养目标、高等医学教育教学改革的需要和医药卫生行业人才的需求，借鉴国内外医学教育教学的经验和成果，不断创新编写思路和编写模式，不断完善表现形式和内容，不断提升编写水平和质量，已逐渐将每一部教材打造成了学科精品教材，使第八轮全套教材更加成熟、完善和科学，从而构建了适合“5+3”为主体的医学教育综合改革需要和卓越临床医师培养需求的教材体系，推动了适合中国国情的五年制本科临床医学专业课程体系的建设。

本次修订和编写特点如下：

1. 教材编写修订工作是在教育部、卫生部的领导和支持下，按照"5+3"为主体的临床医学教育综合改革的时间表、路线图和施工图进行顶层设计，由全国高等医药教材建设研究会规划，全国临床医学专业教材评审委员会审定，院士、专家把关，全国各医学院校知名专家、教授编写，人民卫生出版社高质量精品出版。

2. 教材编写修订工作是根据教育部培养目标、卫生部行业要求、社会用人需求，在全国进行科学调研的基础上，借鉴国内外医学人才培养模式和教材建设经验，充分研究论证本专业人才素质要求、学科体系构成、课程体系设计和教材体系规划后，科学进行的。

3. 在全国广泛、深入调研的基础上，总结和汲取了前七轮教材的编写经验和成果，尤其是对一些不足之处进行了大量的修改和完善，并在充分体现科学性、权威性的基础上，更考虑其全国范围的代表性和适用性。

4. 教材编写修订工作着力进行课程体系的优化改革和教材体系的建设创新——科学整合课程、淡化学科意识、实现整体优化、注重系统科学、保证点面结合。继续坚持"三基、五性、三特定"的教材编写原则，以确保教材质量。

5. 为配合教学改革的需要、减轻学生负担和体现"干细胞"教材特色，全套教材精炼文字、压缩字数，注重提高内容质量，并根据学科需要，采用大16开国际开本、双色或彩色印刷，以提高印装质量和可读性。同时，在每一页都增加了留白，便于学生记录和标记书中重点知识。

6. 为满足教学资源的多样化需求，实现教材系列化、立体化和数字化建设，大部分教材配有配套教材和数字出版的教学资料，并实现了全套教材的网络增值服务，方便老师教学和学生自主学习，实现了数字化资源共享。

第八轮教材共有53种，其中新增2种，即《医患沟通》和《肿瘤学概论》；更名1种，即《急诊医学》更名为《急诊与灾难医学》；合并2种，即《生物化学》与《医学分子生物学》合并为《生物化学与分子生物学》。全套教材均为**"十二五"普通高等教育本科国家级规划教材**（除《肿瘤学概论》外）和**卫生部"十二五"规划教材**，于2013年6月全部出版发行。

本套教材是在我国医学教育综合改革，构建"5+3"为主体的临床医学人才培养体系背景下组织编写的，希望全国各广大院校在使用过程中能够多提供宝贵意见，反馈使用信息，以逐步修改和完善教材内容，提高教材质量，为第九轮教材的修订工作建言献策。

全国高等学校五年制本科临床医学专业

第八轮 教材目录

	书名	版次	主编	副主编
1.	医用高等数学	第6版	主编 张选群	副主编 马建忠 吕 丹 刘春扬
2.	医学物理学	第8版	主编 王 磊 冀 敏	副主编 李晓春 吴明海
3.	基础化学	第8版	主编 魏祖期 刘德育	副主编 李雪华 陈朝军
4.	有机化学	第8版	主编 陆 阳 刘俊义	副主编 叶 玲 邓 健
5.	医学生物学	第8版	主编 傅松滨	副主编 王培林 刘 佳
6.	系统解剖学	第8版	主编 柏树令 应大君	副主编 丁文龙 刘学政 孙晋浩
7.	局部解剖学	第8版	主编 刘树伟 李瑞锡	副主编 张绍祥 羊惠君
8.	组织学与胚胎学	第8版	主编 邹仲之 李继承	副主编 曾园山 周 莉
9.	生物化学与分子生物学	第8版	主编 查锡良 药立波	副主编 周春燕 冯作化 方定志 何凤田
10.	生理学	第8版	主编 朱大年 王庭槐	副主编 罗自强 管又飞 金学隆
11.	医学微生物学	第8版	主编 李 凡 徐志凯	副主编 黄 敏 郭晓奎
12.	人体寄生虫学	第8版	主编 诸欣平 苏 川	副主编 吴忠道 李朝品
13.	医学免疫学	第6版	主编 曹雪涛	副主编 熊思东 姚 智
14.	病理学	第8版	主编 李玉林	副主编 文继舫 唐建武 来茂德 步 宏
15.	病理生理学	第8版	主编 王建枝 殷莲华	副主编 吴立玲 孙连坤 李文斌
16.	药理学	第8版	主编 杨宝峰	副主编 苏定冯
17.	医学心理学	第6版	主编 姚树桥 杨彦春	副主编 杨艳杰 潘 芳 赵旭东
18.	法医学	第6版	主编 王保捷 侯一平	副主编 丛 斌 赵子琴
19.	诊断学	第8版	主编 万学红 卢雪峰	副主编 刘成玉 胡申江 康熙雄 杨 炯
20.	医学影像学	第7版	主编 白人驹 徐 克	副主编 韩 萍 龚启勇 张雪林 王 滨
21.	内科学	第8版	主编 葛均波 徐永健	副主编 梅长林 唐承薇 王 辰 周 晋
22.	外科学	第8版	主编 陈孝平 汪建平	副主编 秦新裕 刘玉村 张英泽
23.	妇产科学	第8版	主编 谢 幸 苟文丽	副主编 林仲秋 狄 文 马 丁 孔北华
24.	儿科学	第8版	主编 王卫平	副主编 毛 萌 李廷玉 申昆玲 常立文
25.	神经病学	第7版	主编 贾建平 陈生弟	副主编 崔丽英 王 伟
26.	精神病学	第7版	主编 郝 伟 于 欣	副主编 许 毅 吴爱勤 李 涛 刘金同

27.	传染病学	第 8 版	主编 李兰娟 任 红	副主编 高志良 牛俊奇
28.	眼科学	第 8 版	主编 赵堪兴 杨培增	副主编 瞿 佳 姚 克
29.	耳鼻咽喉头颈外科学	第 8 版	主编 田勇泉	副主编 韩东一 迟放鲁 孙爱华
30.	口腔科学	第 8 版	主编 张志愿 俞光岩	副主编 凌均棨 杨丕山
31.	皮肤性病学	第 8 版	主编 张学军	副主编 陆洪光 高兴华
32.	核医学	第 8 版	主编 李少林 王荣福	副主编 张永学 匡安仁
33.	流行病学	第 8 版	主编 沈洪兵 齐秀英	副主编 刘 民 叶冬青
34.	卫生学	第 8 版	主编 朱启星	副主编 牛 侨 吴小南
35.	预防医学	第 6 版	主编 傅 华	副主编 段广才 黄国伟
36.	中医学	第 8 版	主编 高鹏翔	副主编 卜 平 陈金水 陈利国
37.	医学计算机应用	第 5 版	主编 袁同山 阳小华	副主编 白宝钢
38.	体育	第 5 版	主编 裴海泓	副主编 程 鹏
39.	医学细胞生物学	第 5 版	主编 陈誉华	副主编 杨 恬 刘艳平
40.	医学遗传学	第 6 版	主编 左 伋	副主编 顾鸣敏 张咸宁
41.	临床药理学	第 5 版	主编 李 俊	副主编 刘克辛 袁 洪
42.	医学统计学	第 6 版	主编 李 康 贺 佳	副主编 杨土保 马 骏
43.	医学伦理学	第 4 版	主编 孙福川 王明旭	副主编 陈晓阳 宫福清
44.	临床流行病学与循证医学	第 4 版	主编 刘续宝 王素萍	副主编 孙业桓 时景璞
45.	康复医学	第 5 版	主编 黄晓琳 燕铁斌	副主编 王宁华 励建安
46.	医学文献检索与论文写作	第 4 版	主编 郭继军	副主编 马 路 张 帆
47.	卫生法	第 4 版	主编 汪建荣	副主编 达庆东 田 侃
48.	医学导论	第 4 版	主编 马建辉 闻德亮	副主编 肖海鹏 郭永松 曹德品
49.	全科医学概论	第 4 版	主编 祝墡珠	副主编 胡传来 路孝琴
50.	麻醉学	第 3 版	主编 杨拔贤 李文志	副主编 刘 进 姚尚龙 郭曲练 邓小明
51.	急诊与灾难医学	第 2 版	主编 沈 洪 刘中民	副主编 王育珊 周荣斌 于学忠
52.	医患沟通		主编 王锦帆 尹 梅	副主编 唐宏宇 赵明杰
53.	肿瘤学概论		主编 王冠军 赫 捷	副主编 张清媛 李 薇 周云峰

第六届全国高等学校五年制本科临床医学专业评审委员会名单

顾　　问

沈晓明　王德炳　刘德培　吴孟超　刘允怡

主任委员

陈灏珠　钟南山

副主任委员

王卫平　杨宝峰　龚非力　柯　杨　石应康　郑树森

委　　员（以姓氏笔画为序）

王　滨　王冠军　王家良　王鸿利　文历阳　文民刚　文继舫
孔北华　田勇泉　白　波　白人驹　冯友梅　吕兆丰　朱明德
刘吉成　闫剑群　李玉林　步　宏　吴在德　吴肇汉　汪建平
沈　悌　陆再英　郎景和　赵　群　赵玉沛　南登崑　柏树令
曹雪涛　崔慧先　葛均波　曾因明　曾晓荣　雷　寒　瞿　佳

序言

高水平、高质量的医学教育既是办好人民满意教育的重要组成部分，也是医疗卫生事业改革发展的重要支撑。随着我国医药卫生体制改革的不断深入，对高等医学教育改革也提出了更高的要求。如何培养适应国家需要、人民满意的高质量、高水平医学人才是当前医学教育的首要任务。为此，在“十二五”开局之年，教育部和卫生部共同组织实施了医学教育综合改革。

医学教育综合改革要求我们深入贯彻落实教育规划纲要和医药卫生体制改革的意见，遵循医学教育规律，以改革创新为动力，着力于医学教育发展与医药卫生事业发展的紧密结合，着力于人才培养模式和体制、机制的重点突破，着力于医学生职业道德和临床实践能力的显著提升，着力于医学教育质量保障体系的明显加强，从而全面提高医学人才培养质量，为发展医药卫生事业和提高人民健康水平提供坚实的人才保障。

教材建设在提高人才培养质量中发挥着重要的基础性作用，对此教育部一直高度重视，要求以教材建设为抓手，推动医学课程和教学方法改革。一本好的教材，给医学生以正确的引导，给临床医生以正确的指导。人民卫生出版社作为国家级优秀出版单位，承担了大量教材的规划和出版工作，形成了课程种类齐全、学科体系合理、配套服务全面的教材出版模式。尤其是在以吴阶平、裘法祖、吴孟超、陈灏珠等院士为代表的老一辈医学大家的付出和带领下，在一大批医学教育精英的努力和参与下，其出版的五年制本科临床医学专业规划教材为我国医学界培养了一代又一代优秀的医药学人才，为推动我国医疗卫生事业的改革和发展做出了巨大的历史贡献。

此次第八轮五年制本科临床医学专业规划教材的修订工作是在贯彻党的十八大关于“深化教育领域综合改革”精神的背景下，在落实卫生部、教育部联合下发的《关于实施临床医学教育综合改革的若干意见》的基础上启动的。修订工作贯穿了医学教育综合改革的要求，特别是注重将医德教育贯穿于医学教育的全过程，增加了《医患沟通》一书，同时强化临床实践教学，配套编写了相关的实践指导，以提高医学生的临床实践能力。

我们相信，在教育、卫生系统的通力合作下，在广大医学教育工作者的大力支持和参与下，第八轮五年制本科临床医学专业规划教材的修订出版对推动医学教育综合改革，提高医学人才培养质量将产生积极的推动作用。

教育部部长助理

林蕙青

2013 年 3 月



王家良

男，1933年9月出生于湖北省宜都市。1960年毕业于原四川医学院医学系（现四川大学华西临床医学院），1982年留学加拿大 McMaster 大学，获理学硕士学位。华西医院内科及临床流行病学教授，曾任大内科副主任和临床流行病学教研室及国际临床流行病学网（INCLEN）二级资源与培训中心主任。系我国临床流行病学及循证医学的主要创始人和奠基者。首编与出版了我国《临床流行病学》和《循证医学》专著；主编了国家重点医学教材《临床流行病学》和《循证医学》（长学制、五年制），还承担了主审长学制教材《临床流行病学》的任务等。先后获国家级一等教学成果奖、教育部二等优秀教材奖、全国医学教材一等奖以及全国优秀图书二等奖，为我国临床流行病学和循证医学的学科系统理论及教材建设奠定了良好的基础。

在国际上先后被聘为国际临床流行病学网（INCLEN）董事会董事及委员会委员，获得了 INCLEN 特别贡献奖。在国内先后倡议、发起和组建了中国临床流行病学网（ChinaCLEN）和中华医学会临床流行病专业委员会并被选为首任主任委员、荣誉主任委员；先后获国家人事部授予的"国家优秀留学归国人员"、卫生部授予的"有突出贡献的中青年专家"等荣誉称号，以及国务院首批"有突出贡献专家津贴"等。



王滨有

男，1942年6月出生于黑龙江省哈尔滨市。现为哈尔滨医科大学公共卫生学院流行病学教授、博士生导师，黑龙江省重点学科流行病学学科带头人。曾先后留学于加拿大多伦多大学、加拿大卡尔加里大学和美国贝勒医学院。曾兼任第四届国务院学位委员会学科评议组成员、省学位委员会委员、中华预防医学会流行病学会副主任委员、中华医学会临床流行病学会副主任委员、中华流行病学杂志副总编辑、全国继续医学教育委员会学科组成员、江苏大学客座教授、山东大学兼职教授、安徽医科大学名誉教授，加拿大多伦多大学客座教授。

从事教学工作46年。享受国务院政府特殊津贴。曾先后荣获全国优秀科技工作者，黑龙江省第三届高等学校教学名师奖，中华预防医学会"公共卫生与预防医学发展贡献奖"，中华预防医学会流行病学分会"中国流行病学杰出贡献奖"，黑龙江省"留学人员报国奖"，哈尔滨医科大学教学名师和哈尔滨医科大学十佳中年教师等多项奖励。先后主持承担国家自然科学基金项目、CMB项目、国际协作项目及卫生部、省科委重大项目、省自然科学基金项目等多项科研项目。曾先后获卫生部、省政府、省卫生厅、省教委科技进步奖多项。作为主编、副主编、编委出版有关专著、教材60余部，于国内外杂志公开发表学术论文170余篇。

主编简介



刘续宝

男，1961年3月出生于四川省成都市。1983年毕业于华西医科大学（现合并于四川大学），获医学学士学位，后留校从事外科临床工作。相继在华西医科大学完成研究生学习，获医学硕士和博士学位。20世纪90年代在菲律宾大学医学院完成临床流行病学理学硕士学位学习和美国印第安纳大学医学院的博士后研究。

现为四川大学华西医院肝胆胰外科和临床流行病学教研室教授、博士研究生导师。兼任中国外科医师协会第一届和第二届常务委员、中华外科学会胰腺学组委员、中国抗癌协会胰腺癌专业委员会副主任委员、中华临床流行病学会第5届和第6届副主任委员、第7届常委，并为四川省医学会临床流行病学专业委员会主任委员、四川省抗癌协会胰腺癌和胃癌专业委员会主任委员，兼《中华外科杂志》和《中华内分泌外科杂志》、《中华胰腺病杂志》、《中国普外基础与临床杂志》、《中国普通外科杂志》等多个专业杂志编委。主要从事普通外科和临床流行病学的临床工作和教学，侧重于胰腺外科疾病的临床和相关基础研究，在国内、外杂志上发表了100余篇研究论文，主编专著《急性胰腺炎》和《腹部外科手术要点及围手术期处理》，副主编第3版“十一五”规划教材《临床流行病学》及参编《胰腺疾病》和《临床流行病学与循证医学》等多部专著和教材。



王素萍

女，1957年8月出生于山西省太原市。现任山西医科大学公共卫生学院主持工作副院长，教研室主任，教授，博士生导师。山西省重点学科流行病与卫生统计学学科带头人，国际流行病学学会会员，中华医学会临床流行病学分会委员，山西预防医学会副会长，中共山西省委联系的高级专家等。

从事临床流行病学及流行病学教学科研工作30年，主要研究方向为传染病流行病学及环境流行病学，主持5项国家自然科学基金，1项科技部重大专项子课题，多项教育部及其他省部级科研基金，国内外期刊发表论文百余篇，获多项科技奖。乙型肝炎病毒宫内传播机制等研究工作成绩显著。培养硕士研究生、博士研究生及博士后90余名。成功申报流行病与卫生统计学博士学位授权点及公共卫生与预防医学博士点。主编及出版普通高等教育“十一五”国家级规划教材《流行病学》，并已获批主编普通高等教育“十二五”国家级规划教材《流行病学》。获中国流行病学优秀奖，获山西省教学成果二等奖，被山西省教育厅及山西省政府授予山西省普通高等学校教学名师，流行病学山西省本科教育精品课程带头人、临床流行病学山西省研究生教育精品课程带头人，流行病与卫生统计学研究生教育优秀导师团队带头人。



孙业桓

男，博士，1961年10月出生于安徽省寿县。现为安徽医科大学教务处处长，循证医药学中心主任，公共卫生学院流行病与卫生统计学系教授，博士研究生导师，安徽省学术和技术带头人，省级重点学科、国家级精品课程《流行病学》学术带头人之一。

近年获省级科技进步三等奖1项，国家级教学研究成果二等奖，省级教学研究成果一等奖、二等奖各1项；在国内外发表学术论文150余篇；主编、副主编、参编出版学术著作或教材16本。获评2007年安徽省优秀教师，2010年安徽省优秀教学管理工作者。



时景璞

男，教授，博士研究生导师。1956年11月30日出生于辽宁省凤城市。现任中国医科大学附属第一医院临床流行病学与循证医学教研室主任，中华医学会临床流行病学分会常委，辽宁省临床流行病学与循证医学分会主任委员等职。任《中国实用内科杂志》常务编委、《中华流行病学杂志》等编委，*Biomedical and Environmental Sciences*、*Gene*、《中华医学杂志》等审稿专家。

作为副主编及编委参加了十几部各类教材的编写，承担了十几项国家、省部级研究课题，作为负责人及主要参加者，获得国家科技进步二等奖1项，省级科技进步奖4项。近些年在国内外杂志发表论文150多篇。

前　言

临床流行病学是一门新兴的、前沿性的、多学科交叉的临床研究方法学。其学科特点是突出临床，传教医学生以群体观点对疾病发生、发展、诊断、防治及预后等方面进行临床研究的基本理论、基本知识和基本方法，使临床医学生从个体患者的特点，扩大到疾病的群体共性规律的认识，旨在培养医学生创新性思维和临床研究能力。

本教材第 1 版、第 2 版、第 3 版问世后，接受了全国高等医学教育和医学继续教育社会实践的考验，相继荣获"教育部优秀教材二等奖（2002）"和"全国高等医药优秀教材一等奖（2005）"，表明了本教材是一本深受欢迎的优秀教材。

根据"早临床、多临床、反复临床"的临床医学教育改革精神，在这次第 4 版的修订中，进一步向临床前移，扩增了循证医学的内容，将教材更名为《临床流行病学与循证医学》，同时结合本学科的国际进展和我国高等医学教育的实际，充分发挥了新的编委会的集体智慧，对第 4 版的内容作了一些创意性修订，本版次教材共设置 16 章，其中：

1. 鉴于第 3 版教材评价及反馈意见均良好，在坚持"三基"、"五性"的基础上，将继续保留第 3 版教材原有的 11 章，其中部分章节进行了更名：①原第四章更名为"医学研究证据的检索与收集"；②原第五章更名为"医学研究文献的评阅"；③原第六章更名为"医学研究证据的系统评价与 meta- 分析"；④原第七章更名为"临床研究设计的基本要求"，并整合了原第十章"影响研究质量的误差与防止措施"和原第十二章"临床研究中的伦理学"的内容，分别作为一节阐述；⑤原第九章更名为"临床研究的常用设计方案"；⑥原第十七章更名为"决策分析"；⑦原第二十一章更名为"临床医学研究项目申报书的撰写"。

2. 以临床科研设计、测量、评价等三大基本原则为主线，以病因、诊断、治疗、预后等四大临床问题为切入点，系统阐述了临床研究方法、评价与循证医学实践案例，将临床流行病学与循证医学有机结合、体现"证据来源于临床又服务于临床"的宗旨，修订了本版次的第八章至第十一章。

3. 强化了医学生在科学探新与临床实践中，发现问题与解决问题的能力，充实和修订了原第三章，增加了"如何提出和构建临床研究问题及循证医学实践问题"等内容。

4. 新增了三章循证医学的相关内容，分别是第十二章"临床实践指南的制定与评价"、第十三章"循证医学的个体化实践"、第十五章"循证医学自我评价"。

我们共同的愿望是在全国一流教材的基础上，紧跟现代科学发展的步伐，不断地吸收最新最佳证据，与时俱进地丰富本教材的科学内涵，使之永葆一流的教材水平，以更好地满足我国临床医学教育改革及人才培养的要求。在本版的修订中，尽管全体编委尽心尽责，但毫无疑问地会有某些不足甚至谬误，敬希应用本教材的师生和同道，给予批评和指正！

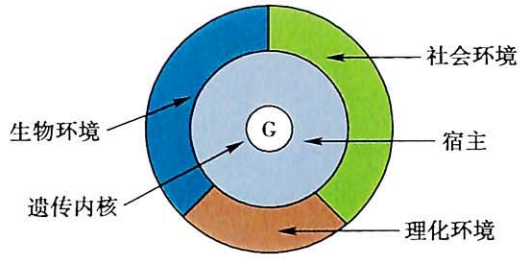
在本教材的修订中，得到了四川大学华西临床医学院/华西医院领导的大力支持，并得到全国高等医药教材建设研究会、卫生部教材办公室及人民卫生出版社的指导和帮助，一并致以衷心的谢意！

在本教材资料图表的编排和辅助教材的编辑以及部分内容的编写、网络增值服务资料的设计制作中，学术秘书康德英和洪旗两位老师做了大量工作，对他们的辛勤劳动和奉献，致以诚挚的感谢！

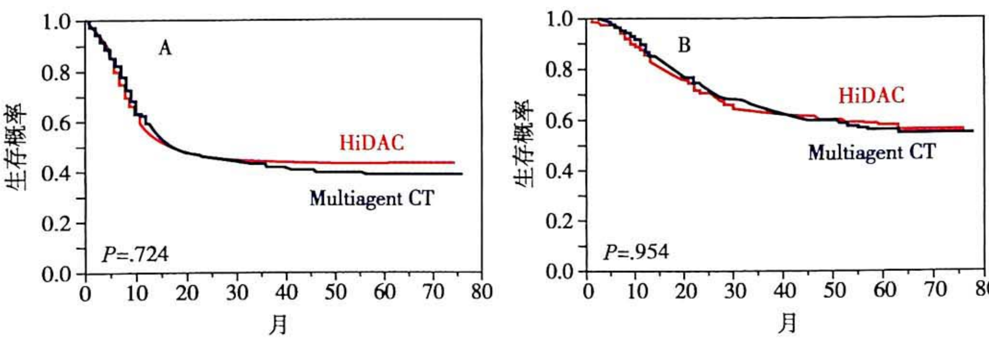
刘续宝　王素萍

2013年1月

目　录

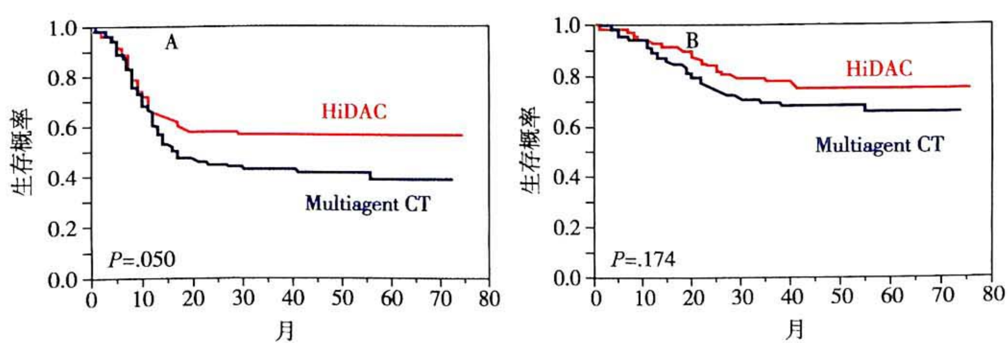


彩图 8-3　病因的轮状模型



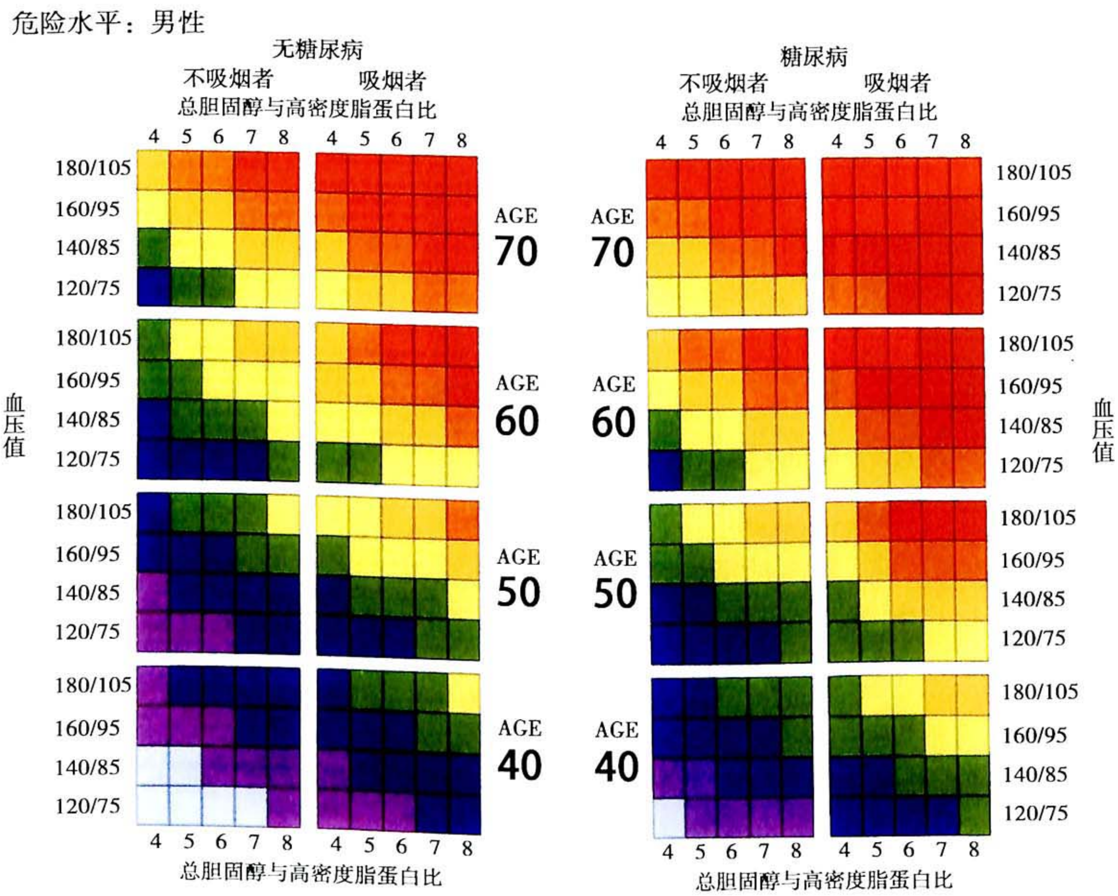
彩图 11-7　两组患者 DFS 和 OS 的生存曲线

A：大剂量 Ara-C（HiDAC）组的 DFS 是 43%，传统化疗（CT）组为 39%（$P=0.724$）；B：HiDAC 组和 CT 组的 5 年 OS 分别是 58% 和 56%（$P=0.954$）



彩图 11-8　染色体预后良好组的 DFS 和 OS

A：HiDAC 和 CT 组的 5 年 DFS 分别为 57% 和 39%（$P=0.050$）；B：HiDAC 和 CT 组的 5 年 OS 分别为 75% 和 66%（$P=0.174$）



彩图 13-2 心血管发病危险因素与治疗效果的评估

第一章 绪 论

临床流行病学(clinical epidemiology)和循证医学(evidence based medicine, EBM)是始于 20 世纪的新兴临床医学基础科学(a basic science for clinical medicine),是临床医生进行临床医学科学研究和指导临床医学实践十分有用的理论与方法学,对促进临床医学的科学发展有着重要的价值。

第一节 历史与现状

20 世纪医学的快速发展促使人们需要采用从定性走向定量、个体研究转向群体研究的科学方法去认识和解决临床面临的、各种不同复杂的医学问题,以便从宏观角度更加科学地指导临床实践。据此,30 年代 John R. Pual 首先提出了临床流行病学的概念,后经几十年的努力,特别是从 60 年代后,著名的内科医师 David L. Sackett、Alvan R. Feinstein 等创造性地将流行病学和医学统计学原理和方法有机地与临床医学的研究和实践结合起来,并进一步拓展到与临床医学相关的卫生经济学和社会医学等领域,极大丰富和发展了临床研究的方法学。在临床研究实践中,提高了对疾病的发生、发展和转归的整体规律的宏观认识,深化了对疾病诊断、治疗和防治方法的科学观,有效地提升了临床医学研究和实践的水平,为现代临床流行病学打下了坚实的基础。

临床流行病学的最初发展和推广,致力于从事临床流行病学研究和工作的临床医师、医学统计学家等功不可没,同时还与国际上致力于医学发展的基金会和医学专业学会的鼎力支持密不可分。在美国洛氏基金会的支持下,1982 年成立了国际临床流行病学网(International Clinical Epidemiology Network, INCLEN)。INCLEN 的宗旨是:"在最可靠的临床证据和最有效地利用卫生资源的基础上,促进临床医学实践,从而致力于改善人类健康。为实现这一目标,本工作网内的各国临床医生、统计学家及社会学家要共同奋斗,以建立和维持最佳的医学研究和医学教育的能力和水平,这些是致力于改善人民健康的最重要的条件"。INCLEN 首先在美国、加拿大和澳大利亚建立了七个一级国际临床流行病学资源和培训中心(The International Clinical Epidemiology Resource and Training Center, CERTC),为全球范围内的著名大学医学院培训了大量的临床流行病学专业人员。此后,又相继在亚洲、非洲、南美洲等地建立了多个二级的区域性的临床流行病学资源和培训中心(Regional Clinical Epidemiology Resource and Training Center),这些资源和培训中心承担了全球和地区性或本国人员培训和研究工作。

我国临床流行病学的起步,始于 1982 年中国改革开放初期,在国际基金会的资助下,我国一批优秀的临床医学科学家分赴美国、加拿大、澳大利亚等一些临床流行病起源、发展和成熟的国家学习和工作。这些学者学成归国后为我国临床流行病学学科的建立、发展和普及做了大量辛勤的工作。以四川大学华西医院王家良教授等为代表的我国一批临床流行病学的奠基人,在学科发展的初期,先是在本单位组织、参与和推动临床流行病学的研究和教学工作,继而在卫生部的大力支持下,1983 年获准在原华西医科大学(现四川大学华西医学中心)、原上海医科大学(现复旦大学上海医学院)和原广州中医学院(现广州中医药大学)建立了三个"设计、衡量、评价(design, measurement and evaluation, DME)"的国家级培训中心(the centre of DME on the medical research),简称 DME 中心。开设了临床医学本科生和研究生的临床流行病学课程,并推动了其他医学院校的临床流行病学教研室(或教研组)的建立和开展相应的教学和研究工作。进而在 1989 年成立了 INCLEN 指导下的中国临床流行病学网(China Clinical Epidemiology Network, ChinaCLEN);由于临床流行病学在中国成功地发展,在 INCLEN 的支持下,分别在当

笔记

时的华西医科大学和上海医科大学成立了两个二级区域性的临床流行病学资源和培训中心(R-CERTC),并在原中国协和医科大学(现北京协和医学院)以及中国中医科学院等七所院校建立国际临床流行病学组(CEU),为国内培养了一批临床流行病学的研究和临床医学的践行人才,促进了临床流行病学在国内高等医学院校的普及和发展。在学科普及的基础上,1993 年正式成立了中华医学会临床流行病学分会,这为本学科在我国的快速发展和迈向国际水平奠定了组织基础。临床流行病学在我国 30 多年的发展,为我国临床医学的快速进步和赶超国际水平作出了重大贡献,也为人类的健康卫生事业作出了卓越贡献。

近数十年来,临床流行病学的蓬勃发展直接推动了各临床学科的科学水平提高,尤其是在加强国际卫生研究能力、对重大国际卫生问题的合作研究、促进发展中国家人民的健康水平、卫生资源的合理利用以及为世界卫生组织和各国政府的卫生决策等,均作出了非凡的贡献或发挥了重要影响。世界卫生组织在其 2004 年的报告中,对临床流行病学的贡献给予了高度评价,指出"临床流行病学学科的建立,对在群体层面上的疾病研究和临床干预作出了巨大贡献。其进展从根本上升华了测量疾病的定量方法,使之在各种群体层面上能够可信地评价干预治疗的结果"。

鉴于临床流行病学的发展促进了临床研究成果的产生,而新的研究成果或称最佳证据(best evidence)应适时地应用于临床实践,方可产生科学与实用价值,从而促进临床医学水平和质量的提高,因此,在 20 世纪的 90 年代,在临床流行病学的基础上,继而催生了"循证医学(evidence based medicine, EBM)"。

循证医学的起源,从哲理上可以追溯到 19 世纪中期。在 20 世纪 80 年代,以 David L. Sackett 为代表的一批临床流行病学家,倡导与催生的现代循证医学,其精华在于运用临床流行病学的原理和方法,检索、阅读和评价相关的医学文献,科学地寻找有效解决这些临床问题的最新科学研究成果,作为临床诊断和治疗决策的证据基础,从而科学地指导临床实践,从而不仅使病人获得最科学、合理、有效的治疗,而且会不断提高医生的水平,使其医疗质量水平永居前沿,故为临床医生终生的一种最佳自我教育的方法。从最初对临床医师如何阅读、评价医学文献培训发展到对发表的众多相关最新研究成果进行二次评价和综合研究的循证医学培训,经过多年的实践,取得了非常好的效果,于 1992 年在 JAMA 杂志上发表了系列的循证医学总结性文献,David L. Sackett 和 B. Haynes 等人的这一工作受到了临床医学各学科的关注。

在 1993 年,国际上专门成立了 Cochrane 协作网(Cochrane collaboration network),在全球范围内接受注册和广泛收集临床随机对照试验的研究结果,在严控质量的基础上,对试验的结果进行系统评价(systematic review, SR)和 meta 分析(meta-analysis),将有价值的临床研究结果推荐给临床专业医师和相关领域的工作者。1995 年,随着 David L. Sackett 受聘于英国牛津大学,成立了全球第一个循证医学中心,继而循证医学专著以及由英国和美国内科医师协会主办的循证医学专业杂志相继出版发行,全面推介国际上经过严格评价的医学成果,以促进这些成果作为指导临床实践的证据,进行临床转化以及应用于临床实践指南(clinical practice guideline, CPG)的制定。

伴随我国临床流行病学的发展,循证医学也获得了快速普及。1996 年在卫生部的支持下,中国循证医学中心和 Cochrane 中国中心在原华西医科大学成立,相继编著与出版了《循证医学》专著,创办了《中国循证医学杂志》;对外在国际上开展了广泛交流和合作,对内组织了多批次国内临床医师和相关专业人员培训,推动了临床医学的进步和发展。

第二节 临床流行病学与循证医学的学科特点

一、临床流行病学和循证医学的概念

笔记

临床流行病学是将现代流行病学及统计学等原理和理论引入临床医学的研究和实践的一

门临床方法学，采用宏观的群体观点和相关的定量化指标，将科学严谨的设计、定量化的测量和严格客观的评价贯穿于临床研究，探索疾病的病因、诊断、治疗和预后的系统性规律，力求避免各种偏倚因素的干扰，确保研究结果的真实性，并对临床医学实践产生重要的循证价值，因此，它重在创造最佳的研究成果，促进临床医学水平的提高。

循证医学是一门临床实践的科学，指医生在自己的临床实践中，针对病人的具体临床问题（难点），谨慎、确切和明智地采用目前最佳的证据对病人的诊治做出科学的决策，从而取得最有效的治疗结果。这样做既能有效地解决病人的临床问题，促进病人康复，同时也会推动临床医疗水平的提高和临床医学的进步。因此，它重在应用最佳的研究成果，解决个体患者具体的临床问题。

可见，临床流行病学与循证医学是理论联系实际、高度统一的临床医学体系，旨在推动临床医学进步与有效地服务于人民的健康事业。

二、临床流行病学和循证医学是以临床医学为主体的多学科合作

临床流行病学和循证医学的学科主体是临床医学，重在临床科研与临床实践。过去，临床医师通常面对的诊治对象是个体，缺乏群体观念，临床研究常常成了个体案例的累加与总结分析，这些经验性的临床研究往往蕴藏了大量的偏倚、混杂和机遇因素，所得出的研究结果或结论往往偏离于客观的真实性。现在，临床医学的研究，以临床为基础，将视角注重于群体观念和定量化的观点，借鉴和采用了大量有关流行病学、卫生统计学、卫生经济学及其他基础医学的原理和方法，创新和发展了新型、科学和实用的临床研究方法（临床流行病学与循证医学），应用这些原理和方法，既有利于创新临床研究，又有助于临床实践，促进临床研究成果转化，服务于临床诊治实践。所以，临床流行病学及其后续的循证医学是一门以临床医学研究为基础的交叉融入了流行病学、卫生统计学、卫生经济学、医学社会学等多门学科的临床医学基础学科。

三、临床流行病学的研究对象是病人及其群体

临床流行病学的研究对象是以医院为基础的病人及其相应的患病群体，这种特定疾病的群体性（specified population of disease）乃为本学科的“流行病学”特征。这种群体性的特征与传统的局限于医院的临床医学有很大的不同，它是从医院的患病病人延伸至社会的特定疾病人群，将医院内特定疾病的病人诊治和社会人群的特定疾病的诊治研究相结合，跨越了医院或社会人群的界限，无疑对疾病的早期发现与防治，以及对疾病发生、发展和转归规律的认识更加全面系统和深入，因此，本学科对临床医学的发展有重要价值和意义。但因其临床基础和基于医院背景下的研究人群，临床流行病学的理论体系、研究范畴、防治实践等多个方面又有别于流行病学。

四、临床流行病学力求研究结果的真实性与可靠性

临床流行病学的精髓还是在于强调：在临床医学研究中采用科学的方法学，强化科学研究的设计和实施，以便从根本上排除各种偏倚、混杂因素对研究结果的影响，确保研究结果是真实情况的反映，即研究结果真实反映了疾病的发病、过程和转归，即研究结果的真实性（validity）和研究结论的可靠性（reliability）。具有真实研究结果和可靠研究结论的成果，对临床的循证诊治决策起到正确的指导作用。

五、循证医学的临床实践基础

循证医学是临床实践活动，当然其基础包括实践活动主体的临床医师、服务主体的病人、遵循的最佳证据、筛检和使用证据的理论与方法以及实践活动的医疗平台。

（一）医师

高素质的临床医师是循证医学实践的主体。良好素质表现为在处理病人的诊治时，既能善

笔记

于利用个人临床技能和丰富的专业知识，同时也善于使用获得的最佳的证据。如果没有优秀的临床技能，再好的证据也难用于病人的诊治；如果医生缺乏识别最新最佳的研究成果的能力，再好的诊治病人的技术和手段，也将不会被采用，因而必然落后于时代。

（二）病人

病人是循证医学实践的服务主体。接受医师的治疗不但是病人出于对自己健康的关注，同时，非常重要的也是出于对医师的信任。任何循证医学实践均需要病人的合作与配合，如果病人缺乏对诊治医师的信任，则依从性就不能获得保证。因此，尊重病人，将病人的利益放在首位，建立良好的医患关系并取得病人的合作是实现循证医学实践的重要基础之一。

（三）最佳证据

最佳证据乃是来源于现代临床医学的研究成果，这些成果经过严谨评价而确认是真实的、有重要临床意义且又有实用价值的，方为最佳证据(best evidence)。最佳证据为实践循证医学的“武器”，是临床上解决病人问题的最新和最佳的手段。而这种证据的获取，则是根据临床面临的具体问题，在全球范围内所有的相应最新研究成果中，应用科学的方法去检索、分析与评价，而获得的最新、最佳的证据成果，并用于临床问题的解决。

（四）临床流行病学的基本知识与方法学

由于分析和评价最佳证据的方法与标准，都源于临床流行病学的基本理论与临床研究质量的评价原则，且诸多的研究成果中，由于研究的设计、方法和研究条件的限制，并非都是最佳的证据。因此，在实践循证医学中使用“最佳证据”的时候，一定要结合临床实际，持批判性的态度。当对某种(些)证据有疑问时，应查其原始来源并用临床流行病学有关理论与方法进行分析和评价，这样方可避免被误导或误用，从而避免对病人的伤害。

（五）医疗平台

循证医学实践都要在具体的医疗平台上实施。地区经济的差异、级别不同的医院，设备条件和医务人员的技术水平各异。某些治疗措施和方法的疗效很好，但需要借助一定的设备或掌握了相当难度的技术的医师作为支持，如果缺乏这些条件，也是难以实现的。因此，循证医学实践应结合具体的医疗平台作为基础，不可盲从。

临床循证医学实践，则首先必须准确地发掘病人存在并需要解决(回答)的临床问题；根据临床问题去检索相关的资料及文献；对收集到的文献进行严格的评价，以找出“最新最佳证据”；针对病人的具体情况，权衡利弊，将最佳证据用以循证医疗决策付诸执行；最后分析与评价最终结果。这就是实践循证医学的“五步曲”(图1-1)。

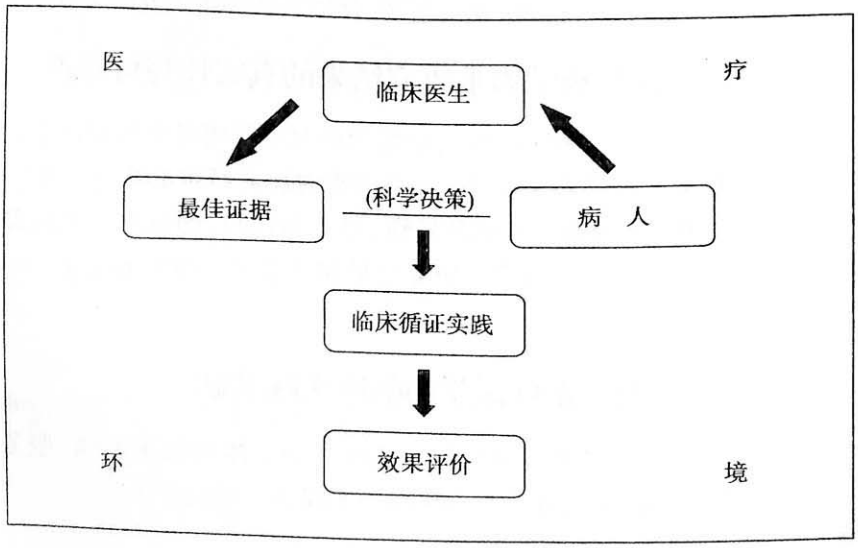


图1-1 临床循证实践示意图

笔记

第三节 临床流行病学与流行病学的关系

流行病学是研究特定人群中的健康相关状况或疾病事件的分布及其决定因素，同时要应用这种研究的结果去解决健康问题[1]。

可见，该学科是定位于人类社会中的特定人群，目的是研究人群的健康相关状况和疾病事件的分布和它们发生的相关因素，并应用研究的结果去有效地改善人民健康并控制相关的疾病。例如：某些急性传染病——天花、霍乱、伤寒、病毒性肝炎、严重急性呼吸综合征(SARS)等在人群中的发生与流行规律以及应用研究出的特异性疫苗和相应措施在人群中进行有效防治，以解决人群中的健康问题。

临床流行病学则是临床学家学习与应用了流行病学的原理与方法，结合了临床医学的实际，研究特定疾病的“群体”分布及其临床特征，包括病因、危险因素、诊断、治疗以及预后等，以期获得最佳研究结果，并且要应用这些具有群体特征性的研究结果，面对个体患者的具体情况，作出科学预测性的防治决策，以解决具体病人的健康问题，可见，它所涉及的是具体的疾病及其在临床方面的诊治。这与流行病学是有所差异的。因而，尽管在解决人类健康或疾病防治等问题上各有侧重，但在学科的理论和方法学上却有着密切的关联。正如临床流行病学的创始人 John R.Paul 早年提出的“临床流行病学”是将流行病学家在人群中用于研究疾病的定量化的概念和临床医学(生)日常用于对个体患者诊治决策之间的紧密融合[2]。

Robert H. Fletcher 在论证临床流行病学与流行病学两者关系时，颇为精辟地认为：临床流行病学本身是源于临床医学和流行病学两大学科，说它是“临床”是因为要回答的是临床问题，并用最佳证据指导临床决策；说它是“流行病学”则指用以回答临床问题的方法，很多是由流行病学家所创用的，同时对个体患者医疗决策的最佳证据往往源于对患病群体的较大样本的研究成果。

当前，我国临床流行病学的发展，在专业队伍中，一部分为临床医生经过正规的临床流行病学专业培训而成为临床流行病学“家”；另一部分为流行病学家通过面向临床医学的临床研究与教学实践而“过渡”成为临床流行病学家。正如前述这种学科彼此交叉和结合，必将有利于科学的发展。

第四节 临床流行病学与循证医学的方法学

临床研究的复杂性源于病人的临床特点，即疾病病程、心理状态和社会经济状况等各异，以及机体不同系统、器官病变间的相互影响，构成了同一疾病的不同患者临床表现的不一致性。而临床研究的基本资料来源是病人的症状及相关病史、体检发现的阳性及重要的阴性体征、实验室及某些特殊检查的结果，这些资料的采集出于多个专业部门和人员，人员的临床专业素质、业务水平并非相同，即研究的实施人员之间也存在认知的不一致性。因此，病史的收集整理是否符合实际，体征是否准确，实验等检查结果是否真实可靠，多种明显因素和隐匿因素(偏倚或 / 和混杂因素)都可能对结果的真实与否产生影响，并直接关系到临床研究的质量。面对病人的复杂临床状况以及多源性的临床资料，如果没有科学的方法来收集、整理和处理，则难以识别及防止各种已知和未知偏倚因素的干扰，那么就很难保证临床研究结果和诊治疾病的科学

---

[1] Epidemiology: The study of the distribution and determinants of health-related states or events in specified populations, and application of this study to control of health problems. IEA.A dictionary of Epidemiology.4th ed.2001: 62

[2] A marriage between quantitative concept used by epidemiologists to study disease in populations and decision-making in the individual case which is the daily fare of clinical medicine. IEA.A Dictionary of Epidemiology.4th ed.2001: 30

笔记

性和可靠性。因此，采用这些研究结果来指导临床的诊治，很可能导致决策失误，使病人接受并非正确或最佳的诊断和治疗。

回顾性的传统临床医学研究，通常是在现有资料基础上进行，由于资料并非为所进行的研究而收集，必然存在一些偏倚和混杂因素，因此，结果的真实性很受影响。即使进行的研究是前瞻性的，如果存在严重的设计缺陷，导致不能有效控制已知和未知的偏倚因素和（或）混杂因素，其研究的结果也必然偏离真实的结果。因此，面对临床医学研究中的复杂情况，临床流行病学的先驱们，创造性地将临床研究资料的特殊性与流行病学和统计学的方法学结合，建立了以设计（design）、测量（measurement）、评价（evaluation）为三大核心的临床科研方法学（DME），应用于复杂的临床医学研究之中。

一、设 计

临床研究的设计主要围绕以下六大要素进行：①疾病病人 / 人群（patients/population）：即研究的疾病和病人人群来自什么地方的医院和人群、采用的诊断标准及有无明确的纳入 / 排除标准、多大的样本量、受试者是否愿意参与试验、依从性如何等都需要明确和说明。②研究设计（design）：是何种方案，设计方案的科学性与实用性如何。③干预措施（intervention）：是否有科学依据，有效性和安全性的前期试验依据，具体的应用方法、疗效与不良反应的指标、是否同时采用其他辅助疗法等。④对照（comparison）的方法：对照组受试对象的入选标准和方法、来源何处、同期或非同期、对照组接受的干预措施是安慰剂或其他有效干预措施等。⑤研究的结果（outcomes）：反映有益（有效）及不良反应（有害）的结果的指标依据，判断的方法以及校正与否等。⑥时间因素（time）：干预措施的疗程需要多长时间，是否能满足药效或不良反应显效的时间要求，即试验研究回答试验组与对照组效应差异的统计学所需的最低限度水平（minimally important difference，MID）所用的时间。对于某些疾病设定的追踪观察时间的长短，需要注意疾病的自然病史的时间规律。“PDICOT”是上述要点英文名词的第一字母的缩写，牢记“PDICOT”，对自己的临床研究和评价阅读的文献将是非常有益的。

（一）设计方案的合理选择

任何临床研究的实施都脱离不了设计方案，一个研究的成功与否，设计方案的质量是非常重要的一环。根据不同性质的临床研究课题和研究条件选择设计方案，不仅要注重方案的科学性，同时也要考虑方案的可行性。通常来讲，在可行性的基础上选择论证强度越强的设计方案，同时避免为了片面追求论证强度而忽略了方案实施的条件，使研究难以在现实条件下进行。表1-1 列出了不同性质研究通常应用的设计方案。表内的“+”数目表示论证强度及可行性，“+”的数目越多则表示论证强度越高或可行性越好。

（二）研究对象的正确诊断

为确保研究对象的可靠性，需要采用临床公认的诊断标准对研究对象做出正确的诊断，确保研究疾病分类的准确性。同时按课题设计的具体要求，确定严格、明确的纳入标准及排除标准，保证研究对象的基线一致性和减少非研究因素对结果的影响。

（三）研究对象的来源与样本含量的估算

临床研究的背景是医院，研究对象也主要是医院就诊的病人。根据研究设计的要求，研究对象可以是收治的系列病人，但对于患病率高的慢性疾病（如高血压的治疗研究），研究对象也可以从患病群体中随机抽样产生。

样本量是临床研究所需要的基本的研究对象（病人）的例数，基于这个样本量的研究结果基本可以代表研究对象总体的客观性质和效应特征，又不至于存在影响研究结果与总体实际情况的误差太大的风险。样本量过小，研究结果可能不能反映总体的实际情况，样本量过大，不仅延长了研究时间，并造成人力资源浪费和成本增高。通常样本量根据具有产生有临床价值的效

笔记

表 1-1 不同性质的研究的备选研究方案

研究的性质	备选方案	论证强度	可行性
病因 / 危险因素 (aetiology/risk factor)	——随机对照试验(randomized controlled trial, RCT)	++++	-
	——队列研究(cohort study)	+++	+++
	——病例对照研究(case-control study)	+	+++
	——描述性研究(descriptive study)*	±	++++
诊断性试验 (diagnostic study)	——金标准方法对照,系列诊断指标评价	++++	+++
防治性研究 (prevention/treatment)	——随机对照试验(RCT)	++++	++
	——交叉试验(cross-over study)	++	++
	——前 - 后对照试验(before-after study)	+	++
	——病例对照研究(case-control study)	±	+++
	——描述性研究(descriptive study)	±	++++
预后研究 (prognosis)	——队列研究(cohort study)	+++	++
	——病例对照研究(case-control study)	+	+++
	——描述性研究(descriptive study)	±	++++
	——疾病的现况调查	±	++++
	——叙述性研究	±	++++

* 含横断面调查(cross-sectional study)

应值作为试验组与对照组的假设性差异水平值,以及容许的低水平概率错误水平(包括总体之间并非存在差异而研究结果错误认为存在差异的Ⅰ型错误和总体之间存在差异而研究结果未能给出的Ⅱ型错误),并依据设计的研究方案的性质,选择相应的样本含量的计算公式进行计算。

(四) 创新性的试验措施

创新性的试验干预措施才具有推动相关医学进步的意义,重复性研究不仅价值有限,并可能造成人力、物力等资源的浪费。同时,试验干预措施还应具备有效性的科学依据,安全性也是需要考虑的重要方面,需要有相关的研究证据,对接受试验措施的对象要有安全性的保障。

(五) 试验观察期的确定

临床试验观察期的时间长度要根据试验终点的设计指标而定,如终点指标是痊愈、死亡、有效、无效等。大多数试验对象达到终点需要的时间即定为试验观察期,这需要考虑疾病本身的自然病程和其他药物治疗反应的临床过程。对于通常药物治疗性临床试验的观察期的确定,宜采用 MID 原则,它的确定要有生物学及临床的试验依据。观察过短易致假阴性结论,过长则导致资源的浪费。

(六) 识别和防止偏倚因素的措施

由于临床研究的复杂性,各种已知和未知的偏倚因素不可避免地存在于研究的过程的各个环节,偏倚因素不仅可能对研究的结果产生扭曲,甚至颠覆研究结果的客观真实性。因此在设计中要估计到研究中可能存在的偏倚因素,建立具体的防止和排除偏倚因素的对策,确保研究结果真实可靠。

(七) 统计学分析方法的应用

根据研究采用的设计方法、采用的有关测试的指标和试验干预措施可能产生的预期结果及其相关资料的类型,正确地使用有关统计学的分析方法。临床研究的资料有定量的、定性的,也有等级的;有的是配对的,有的是非配对的;有的研究结果需要多组比较,也有的仅两组间的结果相比较;有的需要作单因素分析,也有的需要作多因素分析;总之,不同类型的研究资料,应采用不同的统计学方法,予以分析比较和评价。正确选择统计学方法对提高研究质量是非常

笔记

重要的，统计学方法选择错误或不恰当，常可导致对结果的错误分析而得出错误的结论。另外，所涉及的统计分析方法应在研究设计的阶段根据研究试验的检测指标和资料类型就应基本确定，而不是在研究结束后对所获的资料和相关数据，随意选择统计学的方法，这不是严谨、科学的做法。对进行临床研究的临床医师而言，如何选择正确的统计学，可以多倾听统计学专业的人员的建议和意见，并进行讨论，同时善于利用现有的统计学软件。为避免倾向性，将结果的统计学分析交由医学统计专业工作者进行盲法处理是更合理的做法。

（八）科研道德

临床试验的受试对象是病人，受试者享有知情权利。因此，应将研究的目的、参加试验可能的获益和可能发生的问题包括不良反应等，以及研究者的责任和受试者的义务告知受试者或其监护人，让受试者评估、理解和按自己的意愿决定是否参与。不应隐瞒受试者在试验中可能发生的风险或夸大受试者可能获得的利益。并且，在研究的设计和执行的各个环节都要充分考虑和尊重病人的权利，维护病人的利益。

这八个方面是临床研究设计方案时需要重点考虑的，也可以说是方案设计的基本框架。

二、测 量

用某种方法或指标来发现、确定和计量病人接受药物治疗或者其他治疗后的效应，或者用某种方法或指标发现确定和计量人体对某种致病因素产生的反应，这些发现、确定和计量治疗效应或治病效应的方法称为测量，并将测量的结果作为治疗效果或诊断的依据。为了准确地对产生致病或治疗效应的测量，需要注意下列几点：

（一）试验措施的效应性和可度量性

作为试验的致病因素或治疗措施，其本身应该具有产生致病或治疗的效应，且这种效应能客观地确定并被临床及实验室等检查方法及指标度量。

（二）测量方法的灵敏性和特异性

测量方法应具有良好的灵敏性，能够发现致病或治疗效应，并且有良好的特异性对这些效应予以确定。否则，会发生漏诊（测）或误诊（测）的错误。

（三）测量指标的数据和临床意义

测量临床效应的指标，包括定量数据、多项等级数据和定性数据和计数数据。定量的数据如血液生化指标、身高、体重等，多项等级数据如病情的轻、中、重，定性数据死亡、存活和有效、无效等，计数数据包括病死率、并发症发生率、伤残率等。针对众多数据，均需要根据公认的临床标准简化为临床上具有重要价值的也是病人最为关心的定性数据，如有效或无效等，但在表现效应量上，定量数据最能说明问题。同时需要说明的是在测量结果的判断上，重视结果的临床意义和统计学意义的综合分析，如治疗效应量或治疗有效率是否具有临床意义和统计学意义，如果二者皆有，则结果有实用价值，如果二者皆无，则结果没有任何价值。关键的问题是结果只有其一，对临床价值来讲，更看重的是临床意义，单纯具有统计学意义而无临床意义的结果同样也没有实用价值。

三、评 价

临床研究结果是否真实可靠、临床意义和实用价值有多大、研究结果能否适用于临床实践及其适用程度等，都是研究者对自己研究成果的评价需要回答的问题，同时也是临床医生能否采用某一研究成果作为循证医学临床决策时的证据所要予以严格评价的问题。

（一）研究的真实性（validity）

无论研究得到的是阳性的结果还是阴性的结果，都需要对研究的客观测量指标做出科学的解释和结论，确保研究的真实性及其可信程度。

笔记

真实性的评价贯穿于整个研究方案的各个环节，如设计方案的性质和论证强度以及存在的缺陷。具体到研究过程，应着重了解对照组的有无以及设置是否合理、研究对象的诊断标准是否科学可靠、纳入 / 排除标准是否恰当、样本量是否足够、组间重要的基线状况是否可比、有无相关偏倚因素存在以及是否采取了相应的防止或处理的措施、受试者依从性如何、对相应的试验观测指标及资料所采用的整理、统计分析方法是否正确等。

（二）临床重要性（clinical importance）

如果研究结果的真实性良好，对临床更重要的是研究结果的临床意义和价值。具有临床价值的研究结果不仅提高人们对疾病及其治疗的认识，而且可以作为具有指导临床实践的循证医学证据。当然，这些重要性需要具体量化的指标来体现。

量化临床研究结果重要性的指标有多种，主要为事件发生率及其变化，归纳为事件发生率（event rate）（如病死率、生存率、治愈率……）、绝对危险降低率（absolute risk reduction，ARR）、绝对获益增高率（absolute benefit increase，ABI）、相对危险降低率（relative risk reduction，RRR）、相对获益增高率（relative benefit increase，RBI），需治疗多少例患者才能获得一例最佳效果（number needed to treat，NNT）以及需治疗多少例患者才能发现一例不良反应（number needed to harm，NNH），根据它们的具体数据可计算各自的 95% 可信区间。此外，用于诊断性试验评价患者亦有一系列量化指标（见本教材第九章）。这些量化指标都能清楚地表明：试验组与对照组相比所示的具体有效以及不良反应程度，便于临床评价。

传统的临床研究结果，特别是定量化的资料，常常计算组间均值以互相比较，如果 $P<0.05$，则认为组间差异有统计学意义，往往下结论认为某组的效果优于另一组的效果。这里要强调的是组间差值的大小究竟有无临床意义，如果没有临床意义的话，即使具有统计学差异，也无临床应用价值。因此，统计学分析组间差异的显著性并不能完全代表临床的意义。所以，对于任何临床研究的结果，务必要注重临床的价值，如果研究的结果具有临床意义的话，那么必须应用正确的统计学方法对结果进行显著性检验，以评价临床差异的真实程度，即肯定结果的真阳性、真阴性的概率以及检验效能的水平。当某种研究结果既有临床意义，又有统计学的差异显著性时，即能作出肯定性的结论；如仅有临床的意义而统计学差异并不显著时，不能因此而否定临床的价值，此时应计算Ⅱ型错误和检验效能的水平；如果一个研究的结果既无临床意义，又无统计学显著差异，则这种研究的结果应予否定（表 1-2）。

表 1-2　临床意义与统计学意义评价

	临床意义	统计学意义	结论
A	+	+	+
B	−	+	−
C	+	−	+~±
D	−	−	−

临床研究的结果，必要时还应做卫生经济学的评价，计算其成本 - 效果（cost-effectiveness），成本 - 效益（cost-benefit）以及成本 - 效用（cost-utility），并进行比较和评价，分析它们可产出的社会效益及经济效益，以肯定那些成本既低，其效果又佳的研究成果，使之能推广应用。

（三）研究成果的适用性（applicability）

临床研究往往是以解决某种（些）重大疾病的早期正确诊断以及有效防治或改善疾病预后等为其目的的，因此，它们的针对性是颇为明确的。因此，对于研究成果的实用性，要分析它们有无实用价值、有多大的实用价值，利弊比有多大，在什么样的医疗环境和条件下可以采用或推广，宜作实事求是的估价。切不可脱离自己的环境技术条件和病人的实际情况，盲目地接受或推行。

笔记

(四)真实性、重要性和适用性评价在循证医学中的具体应用:系统评价

循证医学的重要目的是利用现有的临床研究的成果作为证据去解决临床问题,但是如何寻找和鉴别成果的科学性、真实性和为临床医师自己诊治病人所应用,临床医师和临床研究工作者必须掌握上述的临床流行病学的设计、测量和评价的研究方法学,同时还需要掌握对大量的相关临床研究信息进行定性合成和定量合成的系统评价的方法学,以适应解决临床问题的需要。

事实上,在信息爆炸的时代,面对发表的浩瀚海洋般的文献资料,临床医师要掌握本专业的新技术、新方法和新进展的研究成果,即使不间断地阅读本专业的文献,也难以全面、准确地把握。同时,即使针对临床的同一个问题,我们通过检索会吃惊地发现,可能会有大量的进行过相关研究而发表的文献,但是结果和结论不尽相同或相近,甚至截然相反。例如,生长抑素和其衍生物对胰腺外科临床常见的术后胰瘘并发症的预防,甚至前瞻性的随机对照研究的结果和结论对其预防胰瘘作用的评估也是不同的。类似的问题在其他研究中也常常见到,这就提出一个问题即临床医师如何科学地面对这些不尽相同或相互矛盾的信息,进行科学、系统的甄别,为临床决策提供高质量的、综合性正确的证据。因此,迫切需要一种科学的方法全面收集相关的文献,并进行严格的评价、分析和合成,去伪存真,力求真实可靠的结论为临床实践服务,这就是循证医学的重要方法之一的系统评价(systematic review)。系统评价的方法可以为临床医师和卫生部门提供了做出决策所需要的大量科学信息。因此,系统评价是真实性、重要性和实用性评价在循证医学中的具体应用。

但是我们应该清楚,系统评价本身是一种方法学,具有双刃剑的性质。如果应用得当、采用方法正确,通过对相关的临床研究成果进行严格的评价、分析和合成,可以达到解决多个研究结论不一致的问题,为临床决策提供正确和科学的证据。如果方法不恰当,二次研究的结论可能提供的是不正确的信息,误导临床的决策过程。如一些做定量合成的 meta 分析,由于收集的相关研究不全面或使用的临床研究的结果不可靠,所做出的二次研究即系统评价的结论难免受到多种偏倚因素的影响而与实际情况不符。因此,系统评价的方法正确与否可直接影响到结论的真实性和可靠性(参见本教材第七章)。

第五节 临床流行病学与循证医学对临床医学的作用与价值

一、临床流行病学为临床医学研究提供科学的方法学、催生高质量的成果

临床流行病学的诞生和发展弥补了传统临床研究方法的缺陷和不足,对促进临床研究水平的进步和产生高质量的研究成果,为临床循证医学实践和提高医疗质量服务。

临床研究的最新成果,特别是高质量的成果,可能成为指导临床实践和防治疾病的重要工具和手段,许多研究成果如今已经作为指导临床诊断、治疗和预防的最佳证据(best evidence),诸如一些重要的临床诊治指南的治疗建议均是以现代的高质量的研究成果作为基础。据世界卫生组织统计,全球每年投入的卫生研究经费逾 700 多亿美元;我国政府在“十一五”和“十二五”国家重大疾病防治研究领域,均大大加强了经费投入,期望能通过加强卫生研究,产出高质量的研究成果以促进防病治病的水平。根据“十一五”国家重大疾病防治攻关课题的质量分析,近 30%~40% 的课题因为研究设计方法学的缺陷严重地影响了研究的质量,是为教训。所以一旦研究课题确立后,应用临床流行病学的理论、原则和方法来规范和指引研究是成功的关键之一。

临床流行病学指导临床研究,设计了一种效应环的原理,首先通过调查了解有关疾病对国家和人民造成的疾病负担程度(burden of disease,BOD),以掌握危害人类健康的重大疾病;从特定群体的角度针对它们发病的病因和危险因素,采取针对性的有效防治措施,并不断地总结其

笔记

效率和效果；通过研究实践，收集有关研究成果予以综合评价和进一步地改进，再付诸实践，最后验证疾病危害人类健康的疾病负担减轻的程度。在这样一个循环效应下，要不断地创造新方法、新经验、新措施、新成果，服务于临床医学。

二、循证医学促进临床实践、提高医疗水平

21世纪的科学技术的快速发展，特别是信息技术、生命科学技术的突飞猛进，推动着医学及人类健康水平的提高。众多新的药物、新的治疗方法、新的诊断技术不断地问世，并涌向临床医学领域。临床医师如何选择这些新的技术、新的方法和药物为病员服务，面临的不仅仅是机遇，也是挑战。因为新的技术和新的方法以及新的治疗药物众多，多数的确可以提高临床诊治水平，促进临床医学水平的进步，使病人获益；少数非但无效，并可能给患者带来严重的不良反应、并发症或使病人延误诊断和治疗；还有一些曾被认为有临床价值且被普遍接受的治疗措施或药物，后经严格的临床研究证明是无效或无益的。因此，临床医生如何从众多新成果中去辨别真伪，以提高自己临床诊治水平，对此，临床流行病学和循证医学提供了对疾病病因、诊断、治疗及预后的一系列评价的原则、标准与方法。

循证医学对临床工作的推进表现在两个方面，一是通过对现有的临床研究成果进行检索、系统评价、meta分析等，获得对指导临床工作真正有益和有用的证据；二是从临床诊断、治疗、预防、预后、不良反应、经济分析、生存质量、卫生技术和医疗质量评估等具体的临床工作中入手，围绕具体问题，进行循证医学实践解决这些问题。

三、临床流行病学与循证医学服务于医学教育、培养高质量的人才

临床流行病学和循证医学的精髓是培养医学生从事临床医学研究和临床医疗实践中的科学的学术，批判性、创造性、科学性思维，掌握临床研究与实践的基本理论、知识和方法以及养成严谨的研究和工作素养，为今后从事临床研究和临床实践打下基础，这对培养高质量、高素质的医学生，使之在毕业后的临床医疗实践中，更好地做好临床医疗和科研工作，对于发展和提高临床医学的水平具有重要的战略意义。

（刘续宝　王家良）

笔记

第二章 临床研究与实践问题的构建

临床医师在临床实践中经常会遇到许多需要解决的实际问题，如针对某个患者应该选择哪项检查才能比较准确地诊断或排除某个疾病？应该进行何种治疗才能更好地改善这个患者的预后？实际上，临床医师对患者的诊治过程就是一个不断提出问题，寻找最佳的解决方法，直至最后解决问题的过程。这样的问题，我们称之为临床实践问题。对患者实践循证医学（evidence based medicine，EBM）的第一步就是找出一个需要解决的实际问题，将其构建成为一个需要回答的临床实践问题。因此，能否找准患者亟须解决的实际问题，对于临床实践循证医学至关重要。

另一方面，临床医师在进行临床研究时需要通过研究回答一个科学问题，如某个疾病的病因或危险因素是什么？影响这个疾病预后的因素有哪些？这样的问题，我们称之为临床研究问题。构建一个恰当的科学问题是开展科学研究的第一步，它对临床研究也同样有十分重要的作用。

第一节 概 述

一、找出临床问题的重要性

（一）实施循证医学的第一步

临床医师应该善于在临床实践中对患者进行观察、发现问题并提出问题。只有提出问题后，临床医师才会带着问题去寻找相关的证据，而后根据可信度最强的证据并结合自己的临床经验、患者的意愿等因素提出恰当、合理的解决方案，最终解决临床问题并使患者获益。因此，找出问题是临床实践循证医学的起点，找不准问题就不可能提出恰如其分的临床实践问题。第一步走不好，必将影响循证医学实践中后续步骤的实施。构建一个恰当的临床实践问题，可以帮助临床医师更好地制定寻找证据的策略，有效地收集、评价证据，最终回答和解决这个临床问题。如果没有相应的证据或收集到的证据科学性不强时，临床医师可以将这个临床实践问题转化成为临床科学问题，从而进一步提出研究计划并加以实施，通过临床研究回答这个临床问题，为临床实践提供证据。

（二）医学发展的需要

有的临床医师认为自己在医学院学到的知识足以回答临床实践中的所有问题而不再需要继续学习，这种想法是错误的。在临床实践中，如果临床医师不能结合患者的实际情况提出有意义的临床问题，不去查阅文献、寻找和评价最新的证据，不对自己的临床知识进行更新，患者就不可能得到最好的诊断和治疗。作为研究者，如果临床医师对临床实践不进行思考和总结，临床医学也不可能取得进步。随着医学进步与发展，针对某个临床问题的答案是会发生变化的，对一个临床问题的认识不断升华才能使之逐渐接近真实。

例如，在诊断学教科书中，黄疸加胆囊无痛性肿大即 Courvoisier 征一直被作为胰头癌的重要体征介绍给学生，但这是不是一个一成不变的定律呢？用该体征诊断胰头癌是否敏感而且特异呢？这是诊断学教科书所没有提到的。回答这个问题需要对瑞士外科医生 Courvoisier 于 1890 年提出的该体征进行重新评价。临床观察发现，该体征在早期胰头癌中并不常见；进一步的研究发现，该体征鉴别良恶性胆道梗阻的价值并不高（阳性似然比 2.6），部分患胆石症的

笔记

患者也可以出现该体征。此外，该体征诊断肝外胆道梗阻性黄疸的敏感度也较低（37%），所以Courvoisier征阴性时也不能排除胰头癌的诊断。从目前临床实践的情况看，该体征对诊断胰头癌的临床价值是较低的，这可能是100多年来医学诊断技术进步的结果。随着医学影像学技术的发展，B超、CT、MR、ERCP和超声内镜在临床实践中得以广泛应用，许多胰头癌患者在出现Courvoisier征之前就已经获得了正确的诊断。

临床医学是一门实践科学，临床医师只有在临床实践中不断地提出问题并通过临床研究回答所提出的问题，才能促使临床医学不断发展进步。

（三）循证医学所赋予的任务

循证医学以解决与患者罹患疾病相关的重要临床问题为核心，因此临床医生在临床实践中必须抓住患者身上关键的临床问题。这些问题往往关系到患者的生死安危，而临床医师如果仅依靠自己的临床知识和经验常常难以准确地回答这些问题。临床医师必须以这些问题为主线，按照循证医学的步骤收集和评价以往的临床研究，选择最佳的证据回答和解决所面临的关键临床问题。可以说，循证医学的第一关键是找准患者身上存在的、临床医师必须回答的关键临床问题。

二、找准临床问题应具备的条件

（一）临床责任心

对患者有高度责任感、关心患者、同情患者的医生，会以患者为中心思考问题，能够在与患者的交流中观察并发现重要的临床问题。

（二）丰富的医学基础知识和临床医学知识

人体各个系统的疾病都有其自身的客观规律，临床医师如果不了解疾病的病因、发病机制和临床表现，不熟悉各种诊断试验和辅助检查的特性和适应证，不了解各种药物的治疗机制、药理作用和可能发生的不良反应，那么当他（她）碰到一个具体的患者时就不可能提出恰当的临床问题。因此，具备系统扎实的医学基础知识和临床知识是找准临床问题的重要基础。

（三）具有一定的人文科学及社会、心理学知识

随着医学模式的改变，人们发现许多疾病的发生与心理、精神因素有关。也有一些疾病的发生虽与此关系不大，如慢性肝病、肿瘤等，但患者对疾病的认识和心态会影响疾病的发展和预后。因此，临床医师不仅要了解患者对疾病的想法、期望及忧虑，也要了解患者的社会经济状况和家庭负担等，这样做有利于发现这方面的临床问题。临床医师也只有掌握一定的人文科学、社会和心理学知识，才能与不同性格的患者顺利地进行沟通，发现患者的心理问题并帮助加以解决，其本身也是疾病治疗的一部分。

（四）扎实的临床基本技能

包括如何接触患者、采集病史、全面的体格检查以及对诊断试验选择和鉴别的能力。临床医师在临床实践中必须弄清患者的病史、认真详细地进行体格检查、掌握相关的辅助检查结果。在此前提下，临床医师才可能找出患者最迫切需要解决的临床问题。

（五）临床综合分析能力和判断能力

临床医师应用已掌握的医学理论知识和临床经验，结合患者的临床资料进行综合分析、逻辑推理，从错综复杂的线索中去伪存真、去粗取精，找出实践中的主要矛盾并加以解决，这反映出临床医师的综合分析能力和判断能力，是准确把握临床问题的必备条件。

上述五点是寻找和提出临床问题的重要必备条件，缺一不可。缺少其中任何一点都不利于找准患者的临床问题。

笔记

第二节 如何构建临床循证问题

一、临床问题的类型

从医学生直至高年资临床医生都可以进行临床研究以及参与循证医学实践活动，但鉴于他们的经验、阅历不一，视角与水平不同，即使在临床实践中面对同一个患者，其发现和提出的临床问题（clinical question）也会大不一样。这些问题大致可归纳为以下三个方面：

（一）一般性问题

主要由初学循证医学的实践者提出，这些问题大多属于“背景性”问题。如：

1. 涉及患者的一般知识性问题 如患者的性别、年龄等。

2. 涉及所患疾病的基本问题 某个具体患者存在的临床问题，如在什么地方、何种环境下发病、何时发病、如何发病及发病的相关因素是什么、主要的临床表现是什么等。

（二）特殊的临床问题

这是临床医师在诊治患者的过程中，特别是充分掌握病史、体征、相关检查结果之后，通过临床综合分析从专业角度找到的问题。一般称为“前景性”问题：

1. 个体患者存在的特殊问题 这些问题不解决则必然影响对患者的正确处理。例如，一个肝硬化患者近期腹水明显增多，对其提出“腹水有无感染”就是一个十分重要的临床问题，如果不能确定是否合并自发性腹膜炎，就无法对其进行正确的治疗。

2. 干预 如何进行干预往往涉及病因、危险因素的暴露、诊断、治疗、预后、患者对病情的理解及依从性等。例如，对一个消化性溃疡患者进行治疗时，必须先针对病因提出问题，包括患者有无幽门螺杆菌感染、有无服用非甾体抗炎药病史、有无应激状态等，这些都影响到治疗方案的选择；对慢性活动性病毒性肝炎患者的治疗，其对病情的理解及依从性就更显得格外重要。

3. 干预措施的选择 同一个疾病的干预措施往往有许多种，每一种措施又各有利弊，这时就存在如何比较抉择干预措施的问题。例如，对恶性肿瘤患者是采取手术还是介入性治疗或放化疗，不仅要根据病情进行分析，将各种干预措施的利弊进行比较，还要考虑到患者的经济能力以及与患者及家属沟通的结果。

4. 干预的最后结局 这是循证医学的实践者追求最佳结局最感兴趣的问题。结局可以是症状、体征的改善或者生存率、死亡率和致残率，使用不同的结局指标找出的问题也不尽相同。

总之，以上四个环节是一个有机整体，循证医学的实践者在发现临床问题时一定要将其牢牢掌握。

（三）患者关心的问题

应该结合患者的具体情况提出问题。例如，罹患同一个疾病但年龄不同的患者所关心的问题是不同的。一项有1012名乳腺癌妇女参加的研究发现，70岁以上的妇女最关心的是癌症治愈和转移的可能性，而小于50岁的妇女关心的是治疗对其性功能的影响；有阳性家族史的妇女最关心的是该病是否有遗传性等。因此，临床医师应针对不同患者的不同情况提出临床需要解决的问题。

二、提出临床问题的形式和方法

（一）提出临床问题的形式

现在许多医学院校已经开展“以问题为中心的学习（problem-based learning，PBL）”，这里的“问题（problem）”是指患者存在的一种症状或体征，如黄疸、腹水、胸痛等。本文所指的“临床问题”是一个可以回答（answerable）的问题（question），如“为这个黄疸患者选择辅助检查时，腹部

笔记

B超和CT哪一项更好?"。尽管"problem"和"question"均可翻译成"问题",但两者的内涵完全不同。下面讨论提出临床问题(question)的形式。

1. 一般性临床问题　一般性问题是与患者或患者罹患疾病有关的一般知识性问题,由以下两部分构成:

(1)问题由疑问词(谁、什么、何时、何地、怎么样、为什么)加动词构成:这些问题的答案在患者入院时常常可以通过询问病史和体格检查得到。例如,每一个主诉都应该包括症状发生的部位、严重程度、数量(如出血量)、起病情况(急性还是慢性、持续性还是进展性)、在什么情况下发生、加重和缓解的因素以及其他关联症状等,了解以往是否发生过与主诉相同的情况,曾经做过哪些检查,是否曾经做过治疗以及如何治疗,对预后有意义或对治疗有影响的既往史,相关疾病的治疗情况等。如以呕血为动词时,就必须弄清楚是谁呕血(患者的性别、年龄等特征)、呕血的性质(颜色、量、次数等)、何时何地发生呕血、呕血时患者有无其他症状以及什么是发生呕血的主因、诱因及其基本病变等。

(2)问题涉及一种疾病或疾病的某一方面:例如,"什么原因引起发热?""急性胰腺炎通常在何时发生并发症?"等。

2. 特殊性的临床问题　在临床实践中,患者与医生都会在诊断、治疗、预后、预防、病因等各个方面提出许多需要解决的临床问题。例如,患者常常会问医生"我患的是什么病?"(诊断问题)、"我为什么会患这个病?"(病因问题)、"这个病应该用什么方法进行治疗?"(治疗问题)、"这个病对我的健康有多大影响,会不会影响我的寿命?"(预后问题)。医生在诊治不同疾病或同一个疾病的不同患者时,提出的问题可能各不相同,归纳起来包括以下几个方面:

(1)患者本次住院或门诊就诊时需要解决的问题以及住院后因病情变化而产生的新问题:临床医师可以对患者发生的每一个症状或体征提出问题。例如,上述呕血的患者在就诊时亟须解决的主要问题是及时止血并弄清呕血的原因;出血停止后患者出现计算力下降、昼夜颠倒、扑翼样震颤等表现,这时患者亟须解决的主要问题就变成了弄清是否出现肝性脑病并对此采取积极的治疗措施。也可以在准确采集病史和体检的过程中发现并提出新的问题。例如,一位中年男性因发现黄疸就诊,临床医师在体检时发现无痛而肿大的胆囊(Courvoisier征),提出的问题即可为"该体征对于梗阻性黄疸和肝细胞性黄疸的鉴别诊断是否有意义",进一步可提问"其对于鉴别肝外梗阻的原因,即结石引起抑或肿瘤引起,是否有帮助"等。

(2)诊断方面的问题:初学者在诊断方面常提出的问题是某个症状、体征或某项辅助检查对某个疾病的诊断价值,即关于诊断试验的敏感度、特异度及似然比等问题;而有多年临床工作经验的医师常提出的问题是某项检查对鉴别诊断的意义。通过询问病史和体格检查,临床医师会有一个诊断假设。为了证实该假设,医生可能会进行一些实验室检查或其他辅助检查来肯定或排除该诊断假设。此时可以针对诊断试验的敏感度、特异度、似然比等提出问题,针对其正确性、可靠性、可接受性、费用及安全性等方面也可提出问题。例如,上述黄疸患者的体征(Courvoisier征阳性)及鉴别诊断所采取的B超、CT、MRCP等辅助检查均可视为诊断试验而提出相应的问题。

再比如,对一个呕血的患者为了寻找出血的部位和原因是否应作急诊胃镜检查?仅此一点就可以提出许多临床问题,如"急诊胃镜检查对诊断上消化道出血的敏感度和特异度如何?""急诊胃镜检查对该患者带来的风险有多大?""对肝硬化患者和非肝硬化患者带来的利和弊有无差别?""急诊胃镜检查的诊断结果是否会影响医生对治疗方案的选择?""有无其他可供选择的诊断措施?"等等。

在选择诊断试验前,还可以对患者的验前概率,即患者在没有进行此项诊断检查前患该病的可能性大小提出问题。例如,上述呕血的患者在没有做急诊胃镜检查前,由食管静脉破裂出血引起呕血的概率是多大?这与患者的基本情况有关。如果这位患者有肝硬化病史,则食管静脉破裂出血的可能性较大;如果以前有过类似的出血史,且胃镜已证实是由食管静脉曲张破裂

笔记

引起，则其验前概率就更高。根据验前概率可提出问题，如“急诊胃镜检查的结果是否影响对该患者所采取的治疗措施？”反之，如果该患者无肝硬化病史但长期服用阿司匹林，则其因食管静脉破裂出血的概率就较小，而急诊胃镜的结果可以用来证实或排除食管静脉破裂出血，其诊断结果对患者治疗方案的选择影响较大。此外，如果已经对患者做出了初步诊断，还可以提出“能否通过某项诊断试验确定该病的严重程度？”“用何种指标来测量随访患者在治疗后的改善情况？”等问题。

（3）治疗方面的问题：如何选择利大于弊的治疗手段？如何从经济学成本 - 效果的角度选择治疗方案？特别有意义的是针对目前的常规疗法提出质疑，如“根据患者目前的病情可以采用什么治疗方法？”“该治疗方法的有效性如何？”“有什么不良反应？”“还有什么可以替代的治疗手段？”“哪一种治疗方法更为有效而花费最少？”“该治疗对患者的生存质量有何影响？”“治疗对患者的预后影响如何？”“患者对治疗的依从性和可接受性如何？”等等。上述胰腺癌患者可以根据疾病的严重度选用外科手术或姑息疗法。如果采用后者，为了减轻黄疸的危害及提高生存质量，可以通过内镜放置内支架引流减轻胆道梗阻。可以进一步提出的问题有“安置金属支架，还是塑料支架，何者为佳？”“为加强支持治疗是否给予患者静脉内营养？”等等。

对于食管静脉破裂出血的肝硬化患者，在出血停止后为了预防再次出血，可以选择外科分流或断流手术治疗、内镜下圈套或注射硬化剂治疗，口服 β- 受体阻滞剂治疗、介入治疗等。这时，必须根据患者的具体情况将这些措施预防再出血的效果、风险、后遗症、疗程以及对生存的影响、费用等逐项列出进行比较选择，在找到证据后临床医师应结合患者的病情提出建议，并在征求患者意见的基础上做出决策。

（4）病因方面的问题：包括“怎样识别疾病的原因及危险因素？”“其发病机制是什么？”等等。例如，对于胰腺癌患者可以提出的病因问题包括“发病的原因是什么？”“有无遗传因素？”“发生胰腺癌的危险因素是什么？”“是否与喝咖啡或饮酒有关？”等等。

（5）预后方面的问题：如何估计患者的临床病程？如何预测可能发生的并发症和最后结局？针对不同的结局测定指标可以提出不同的预后问题。例如，上述预防食管静脉破裂再出血的干预措施对“再出血的发生率”和“患者的生存率”两种预后指标是否有不同的效果？

（二）提出临床问题的方法

临床医师在临床实践中遇到患者存在的问题，想解决却存在知识、能力不足时，就要找准问题并记录下来，通过自己的临床思维对其进行整理和排序，先抓好关键问题并做出如何解决这个（些）问题的策略计划，有的放矢地查阅文献、进行文献评价并选择最佳证据来解决患者的问题。在提出临床问题的方法上，要掌握的是：①涉及的问题一定与患者的诊断、治疗和预后密切相关；②涉及的问题一定与临床实践循证医学、提高医疗水平最为相关；③涉及的问题一定是临床实践中最有用，也是最重要的；④涉及的问题往往也是临床实践循证医学中最为常见的。

如果循证医学的实践者在平时的临床工作中对各种不同患者的临床问题能够做到日积月累并不断用最佳证据予以解决，终身坚持，必成名家，并且也会对循证医学做出巨大贡献。

（三）针对患者的实际情况提出问题

上述两种类型的问题几乎包括了所有需要提出的临床问题（question）。对学生而言，前一类一般性的背景问题可能多一些；对有经验的临床医师来讲，后一类前景问题可能多一些。有些问题不需要进一步查资料就可以回答，但就教学而言，有必要让学生了解如何针对患者的实际情况提出可回答问题。

为了进一步了解患者的情况（一般性问题）和对患者进行进一步处理（特异性问题），在上述工作的基础上可以开出教育处方（educational prescription），即提出患者亟须解决的临床问题（question），以进一步寻找答案。教育处方提出的可回答问题必须十分具体。例如，在讨论一位

笔记

重症胰腺炎患者的治疗措施时，不能提类似于"重症胰腺炎患者该如何治疗？"这样的问题。因为问题太过宽泛，如果据此去检索文献，会有上千篇文章，最终无法归纳总结出回答该问题的答案。因此，提出的问题必须要具体到某一项措施，如有人提出"对重症胰腺炎患者是否需要用静脉内营养或肠内营养"的问题，结合患者实际就可以这样提出问题："全胃肠外营养和肠内营养对于急性重症胰腺炎在降低感染发生率、减少并发症发生率、缩短住院时间和降低死亡率方面，哪一种方法较好？"为了回答这个问题，可以用检索词"parental nutrition or TPN and enteral nutrition and acute severe pancreatitis"检索文献，寻找答案。由此可见，构建的问题必须包括对象（某种疾病、症状或患者）和需要比较的措施，这样查找出来的结果，才能对临床医师决策有所帮助。

（四）为临床科研提出问题

临床实践是临床科研选题的丰富源泉，在日常临床实践中临床医师无时无刻不面对许多关于诊断、治疗、病因、预后等方面的问题，不少诊断方法和治疗手段也有待进一步的科学评价。从临床需要出发提出问题，用可靠的方法进行研究，得到可靠的证据回答所提出的问题，最终指导他人的临床实践。

三、构建临床循证问题的模式

在构建一个具体的临床问题时，可采用国际上常用的 PICO 格式。P 指特定的患病人群（population/participants），I 指干预（intervention/exposure），C 指对照组或另一种可用于比较的干预措施（comparator/control），O 为结局（outcome）。每个临床问题均应由 PICO 四部分构成。下图显示了 3 个临床问题的组成方式：①对于慢性肾功能衰竭尿毒症患者，肾脏移植与血液透析相比在生存率和生存质量上哪种方法更好？② ACEI 与 CCB 合用与单用 CCB 相比，在保护肾功能、降低血压和减少尿蛋白方面是否有更好的效果？③对于频发的尿路感染，长期小剂量应用抗生素是否能预防复发？根据 PICO 确定关键词，便于检索（表 2-1）。

表 2-1　临床问题的组成 PICO

Patient/Population	Intervention/Exposure	Comparator/Control	Outcomes
慢性肾功能衰竭尿毒症	肾脏移植	血液透析	能否改善患者生存质量和生存率
慢性肾病	ACEI＋CCB	CCB	保护肾功能、降低血压、减少尿蛋白
频发尿路感染	长期小剂量抗生素	空白对照	预防复发

总之，要提出一个好的临床问题，需要具备系统扎实的基础与临床专业知识和技能，深入临床实践，善于思考，跟踪本专业的研究进展，学会从患者的角度加以考虑，才能提出和构建良好的临床问题。

第三节　临床研究中如何选题与立题

临床研究的目的在于探索人类疾病发生、发展和转归的规律，提高对疾病的诊断和防治水平，消除或减轻疾病对人体的危害，改善预后，提高人类的健康水平。由于危害人类健康的疾病众多，造成的疾病负担程度各异，对疾病的病因和发病机制的认识程度的差异，诊断与防治措施的效果不一等，造成了临床医学的复杂性。此外，由于当代经济的全球化、科学技术以及信息技术的快速发展，人们对自身健康质量的关注以及对防病治病的要求越来越高。因此，如何选择与确立具有重要意义的课题进行研究，是具有挑战性的重大问题。临床研究应遵循国家

的卫生工作方针与政策，以预防为主、防治结合，改善人民的健康，促进我国临床医学事业的兴旺发达，从而也有利于社会安定和国家经济建设的发展。

一、临床研究的特点

临床医学的研究不同于基础医学的实验性研究，也有异于预防医学的群体健康性研究，其特点表现为以下几个方面：

（一）研究对象是患者

临床研究的对象是患者，患者是有关疾病的“载体”。罹患疾病的患者都饱受疾病的折磨，忍受着不同程度的痛苦，承受着肉体、精神及经济等方面的压力，同时由于各自的病理生理特点、心理情感状态、文化水平及所处自然与社会环境的差异，疾病的临床过程和情况十分复杂，即使同一种疾病的病理损害相似，患者的临床表现却往往有十分显著的个体差异，这就形成了临床研究对象的复杂性。因此，临床研究对象务必要符合有关疾病的诊断标准，同时研究者还要制定恰当的纳入与排除标准，使研究对象具有相对匀质性，从而有利于研究的可比性。

（二）干预措施要安全有效

任何开展临床试验的治疗药物或治疗措施包括诊断性措施，一定要有临床试验前的科学依据，证明对患者具有安全性以及对疾病诊治具有有效性，方可进行立题研究。例如，新药的临床治疗试验，一定要有药物化学成分、药效学依据、病理与毒理、药物代谢动力学等可靠的基础医学实验研究以及一期临床试验证据，证明其安全有效且符合相关的法律法规方可立题进行正式的临床试验。因此，对于缺乏科学依据且无疗效证据的药物或“措施”是不允许用患者做试验研究的。为了对比新的试验措施或药物的疗效，需要设立对照组，而对照组的研究措施也要保证患者的安全性。安慰剂应根据治疗性试验的具体情况和要求使用。

（三）医院是主要的研究场所

临床研究主要是在医院范围内进行，但是当涉及病因或危险因素的致病效应、疾病的早期诊断以及疾病的早期治疗或一、二级预防时，则要面向社区筛选具有临床前期或仅有早期轻微临床表现而尚能正常工作的患病群体进行研究。这种包括医院内和医院外患者的综合性群体研究，具有临床与流行病学研究的双重属性，也是临床研究的一个发展方向。

（四）遵循医学伦理的要求

按照世界医学协会关于人体临床试验的赫尔辛基宣言要求，凡是以人体为研究对象的临床研究，所使用的试验药品或措施，都必须具有充分的科学依据，要安全、有效，保证无损于患者的利益。对于接受试验的患者，要明确地解释接受治疗或试验措施的目的、意义及可能发生的副作用。坚持自愿的原则，尊重患者的人格，不欺瞒研究对象。如果患者接受试验，则需要签署知情同意书，以防日后发生争议。即使在试验进程中，患者也有退出的权利。因此，任何涉及以患者为研究对象的临床试验，在正式立题研究时要高度注意医学伦理问题，必须向有关机构的伦理委员会申请，接受伦理委员会的审查，通过后方可进行研究。

二、选题与立题的原则

尽管国家各级机构确定的科研课题投标指南为研究者提供了研究方向和可选课题，研究人员仍然需要掌握临床研究选题、立题的原则，方能充分发挥科研人员的积极性和创造性。

（一）选择重大疾病进行研究

从疾病负担的调查与分析中，选择高发病率、高致残率、潜在减寿年数（potential years of life lost，PYLL）或伤残调整寿命年（disability adjusted life year，DALY）损失大的、造成疾病负担重、对社会安定影响大、波及地域范围广的疾病为重点课题进行研究，如恶性肿瘤、脑血管疾病、呼吸系统疾病以及心血管疾病、急性传染病、人兽共患病、新生儿疾病等疾病负担十分突出的国

笔记

家确定的重点防治疾病。

人体暴露于致病因素后具有不同的易感性和抵抗力，因此应将高危人群列为重点防治研究的对象。例如：吸烟者多为肺癌发病的高危人群；乙型肝炎病毒感染致慢性肝炎者为肝癌发病的高危人群；高血压、糖尿病、高脂血症等患者为心脑血管病的高危人群；环境、水源与食物污染严重的地区，往往是肿瘤高发区和高危人群所在的地区等。

随着社会进步和医学发展，疾病谱、病原生物体及其疾病负担在不断地发生变化，而人们的认识总是落后于不断变化的客观现实，这就要求研究者不停顿地去研究与探索。相应地，研究重点的选择也必然会发生变动。例如，在新中国成立初期，鼠疫、霍乱和天花等急性、烈性传染病疾病负担非常严重，是当时防治研究的重点。当这些疾病被控制或消灭后，工作则转为以监控为主。随着经济发展和人民生活水平的提高、生活节奏的加快、生态环境的变化，肿瘤和心脑血管病等非传染性疾病逐渐上升为我国主要的健康问题，这时临床研究的选题就应随着国家的人民疾病负担状况的改变而与之相适应。此外，我国幅员辽阔、人口众多，各地区的自然、社会和经济环境有着显著的差异，疾病谱及疾病的负担也存在地区性的差异。因此，防病治病的选题也就有着地方特点，各个地区都有相应的研究重点。例如，南方有关省区的血吸虫病、广东地区的鼻咽癌、河南林县及四川盐亭地区的食管癌等均为所在地区的研究重点。

国家重大疾病防治研究课题的选择与决策除了依靠专家的积极性外，根据一些先进的国际经验，还应吸收有关卫生医疗人员、行政管理人员和社区的人民代表参与，共同讨论，以听取各方面的意见和建议，从科学性和防病治病的现实需要性等多个维度进行综合分析，方能做出符合中国国情的防病研究的科学决策。

（二）研究问题要具体、明确

选择研究课题，首要的是确定研究的焦点，即具体和明确地提出期望要解决的问题。例如，艾滋病的防治研究被选定为国家的重点课题，它所涉及的问题十分广泛，如病因及发病的危险因素、诊断、治疗、预后及预防研究等。从临床研究的角度出发，研究的重点应当是探索有效的防治措施和改善患者预后的方法，这是目前迫切需要解决的实际问题。任何具体的研究要想一揽子地去研究和解决全部问题，往往是不现实的。因此，一定要根据自己的基础、专业特长、兴趣爱好、工作条件、技术力量和经济支持等，实事求是地选择某一个关键问题去重点研究，绝不能在一个研究中企图去解决多个问题，否则极易造成困难的局面，达不到预期的目的。

如果属于综合性的重大课题，由于涉及的内容和研究方法不同，可以根据研究的具体内容，以主课题为中心设立若干分课题（子课题）开展研究。分课题是主课题的一部分，与主课题密切相连。各个分课题的有机综合就构成了总课题的研究整体。

（三）创新性

研究课题的选择，要有自主创新性，有新的见解和新的特色，而不是跟着别人去盲目地重复或赶“时髦”。课题涉及的内容，应力求居于该领域的研究前沿并具有领先水平，或者在过去研究的基础上有所突破和改进，如采用现代科学技术，在分子生物学水平上研究候选基因、基因多态性与疾病发生的联系，应用蛋白质组学技术探讨靶向药物有可能取得比以往组织、细胞水平的研究更为精确的结果。即使对于重大疾病防治措施的研究，也应剖析既往或现存有效干预的最佳水平，寄希望于有新的突破，否则就会造成人力、物力和财力的损失和浪费，且研究本身也往往没有什么意义。

（四）公正性

临床研究应面向广大人民群众亟待解决的健康问题。中国作为世界上最大的发展中国家，国内不同地区的经济发展水平存在着巨大差异，社会贫富差距也颇为显著，城乡医疗卫生和保健事业的水平也很不同。因此，临床研究课题的选择，应该结合我国的国情、地方特点及广大人民（不分贫富）对促进健康和有效防治疾病的要求进行考虑，以充分体现临床研究的社会公正性。

笔记

另一方面，国家和地方的卫生研究经费投入是有限的，人力资源与水平也受某些客观因素的限制。因此在选择临床研究课题的时候，应该考虑将有限的资源投入到最需要解决的、关系大多数人民健康的方面，以体现卫生研究资源利用的社会公正性。

（五）选择足够的研究对象

对疾病病因及危险因素的研究，要有足够的接触和不接触暴露因素的观察对象；诊断及防治试验也要计算研究所需求的最低样本数量。在选题的时候，一定要根据实际情况充分考虑研究对象的来源和数量。如果一个地区或单位在规定期间内有足够的研究对象来参与研究，则该研究可行，否则需要考虑组织具有相同或相似水平的地区、单位进行多中心协作，以保证该研究对样本量的需求。临床试验所采用的新措施或新药物都有研发时间限制，期望尽可能快地验证其临床效果和应用价值。如果因研究对象来源不足而导致试验时间延长，就有可能被更新的研究药物或措施所取代，造成原试验研究失去时效性甚至原有的价值。如果研究对象十分充足，则可视具体情况从特定人群或医院患者群体中按设计要求随机抽样获得最终进入试验的研究对象。

（六）合理的研究设计方案

根据立题研究的课题性质，抉择科学性及可行性良好的设计方案是保障研究课题获得成功的关键之一。

（七）干预措施要安全有效

选择任何干预措施和诊断方法进行研究，都要确保其安全、有效。即使对目前尚无有效治疗的疾病进行治疗性试验，该试验药物或措施也要有科学依据，特别是药效学的实验依据。研究要有先进性和创新性，否则就没有研究的必要。

（八）执行研究措施的可行性

选题的时候，对拟采用的研究措施一定要考虑执行的可行性。

1. 仪器设备及实验室条件　试验研究的措施及试验结果，都需要相应的仪器设备进行测试。例如：涉及抗生素或化学治疗药物的耐药性基因或肿瘤病因学的研究、有关抑癌或致癌基因的研究都需要具备分子生物学实验的仪器设备和实验条件，否则即使研究的问题选得好，客观上却不可行。

2. 要有执行研究课题的配套人才　要有一支具备科研设计、实施和观察、资料收集、统计分析及研究总体指导和管理等技术力量的较强专业队伍，这是保证选题立题后能够执行的重要条件。

3. 干预措施和测试的指标一定不能过繁　采取的试验干预措施或药物在剂型、用法、用药的途径等方面应力求简单易行。测试效应的指标不应过多，在保证准确可靠的前提下选用灵敏性和特异性好的指标，做到少而精既可以保证执行的可行性，也可以减少假阳性或假阴性偏倚的影响，避免误导研究的结果和结论。

4. 注意社会、文化和宗教等的可接受性　临床试验主要是以患者作为研究对象，因此要注意社会习俗、文化素养以及宗教信仰等因素及临床试验措施的可接受性，否则就可能产生负面影响。

5. 注意研究对象的依从性　即使一种很好的防病治病措施，如果价格过贵或应用方法过繁或对患者缺少足够的解释及关爱，则研究对象的依从性往往无法得以保证，研究者在研究课题立题的时候一定要充分注意这一点。

（九）经费支持

在选题和立题时，研究者务必要考虑经费的来源与保障，实事求是地预算研究经费，保证仪器设备、实验试剂、药品开支、劳务报酬和学术交流等必需的费用。不切实际（过高或过低）地预算经费是不利于所立课题的正常运作的。

笔记

（十）伦理学原则

任何临床研究课题的立题，务必要符合伦理学的原则，遵守国际公认的赫尔辛基宣言的精神。按照国际惯例，临床试验的研究课题必须经过有关机构的科研伦理委员会审查通过后方有资格立题研究。

（十一）预测研究成果的价值

研究者要根据科学的依据对选择和立题研究的课题可能获得的科学成果、防病治病的效果和价值进行实事求是的预测，估计其成本 - 效果、是否会被推广应用以及可能产生的社会效益及经济效益等。这些因素及指标对于选题申报具有很重要的参考价值。

三、临床选题和立题的程序

（一）研究问题的提出与选题

临床研究问题的提出，离不开防病治病的具体实践。当在临床医学实践中发现了某种不易解释的临床现象或某种疾病严重地威胁人类的健康时，必然促使人们要去探索其病因及发病的危险因素，弄清楚诊断，为有效地防治提供科学的依据；当某些疾病对患者造成巨大的负担而缺乏有效治疗对策的时候，就会迫使临床学家和基础医学学科及药学专家们去研究发掘有效的防治药物与措施，以挽救患者；对某些难治的、预后不良的急性或慢性疾病，往往会激励临床学家去研究一些改善预后的措施以减轻病残，降低病死率和改善患者的生存质量。人们的认识会在这个过程中逐步地深化和完善，从而找准研究的课题和解决的重点。

新问题的出现，成为人们去研究和解决的动机。从多种认识中去比较分析，去粗存精、弃伪存真，以选出最重要、最能影响全局的问题去研究，这就是从提出的问题中进行选题研究的主要过程。

为了有助于对选题的理解，这里以艾滋病的研究为例。众所周知，艾滋病是 20 世纪末威胁全人类健康的重大问题之一，据 2003 年底 WHO 报告艾滋病（HIV）感染者，全球达 4000 余万人，发病者达 500 万人，死亡者达 300 余万人。当今世界投入研究的资源是十分惊人的。从临床病例的发现，对患者的病例 - 对照研究和流行病学调查，弄清了该病的传播途径和发病的危险因素。在此基础上，病毒学和免疫病学家的研究，发现了致病的人类免疫缺陷病毒（HIV），导致患者的免疫功能严重损害，成为继发感染和致瘤的原因，相继为艾滋病的特异性诊断提供了有力的武器；发病机制的阐明对临床病征的解释和治疗提供了科学的依据。尽管如此，临床却缺乏有效的治疗手段。生物医学家们正全力地研究治疗的药物，以期挽救患者生命并改善预后。在预防研究方面，疫苗的研制与应用、针对发病危险因素的干预难度很大，涉及社会经济、精神文明和道德行为的方方面面。诸多未解决的问题，需要医学家及社会学家从中选择研究，一一突破和解决。这就是对艾滋病的研究及其需要解决的重点问题。

又如 2002 年底至 2003 年初从我国发生的严重急性呼吸综合征（SARS），导致国内及国际上部分地区的传播，经 WHO 组织国际合作研究确定病原为变异的冠状病毒。尽管全球的病例仅 8000 多例，但其病死率高（5%~10%），影响甚大。当前重点是研究该病的特异性的早期诊断方法以及有效的防治措施，改善患者的预后，这些是亟待解决的重大课题。因此，国际卫生研究发展委员会（COHRED）提出卫生研究的选题原则，强调两点：一为国家重大的疾病；二为这种疾病对全球人民的健康具有重要的影响。对具有这两方面特点的疾病应重点选题进行研究。因此，这些原则与事例可作为研究者选题的参考。

（二）充分掌握现有的科技信息

人们的认识水平和临床经验是十分有限的。当出现陌生或不易解释的临床现象，或者发生了某种未知疾病，临床医学家必须回答这些挑战性的问题。这些问题，可能是医学领域里的未知现象，或者是不太清楚的事件，也可能是自己不知而他人已知的事件。因此，除了应用自己

笔记

的聪明才智之外，要带着问题去认真地检索医学文献，充分掌握科技信息资料，认真进行分析，以了解被研究问题的历史、现状以及尚待探讨的关键问题。抓住了矛盾与问题，则应进一步地追寻采用什么样的研究方法去研究解决这类具有挑战性的问题。

在查阅文献的时候，一定要采用临床流行病学/循证医学的原则和方法，进行严格的科学分析和评价，吸收有充分科学依据、真实性和水平高的文献精华；抛弃有偏倚因素影响的、缺乏科学依据的文献。这样，对研究课题的选择才不至于发生误导。这里要十分强调：好的文献并不一定都是阳性的结果，阴性的研究结果并不意味着是质量差的文献，标准取决于文献的真实性。

为研究选题而查阅文献，要充分地应用现代科学的检索手段和方法，针对研究的课题，尽可能收集高度相关、高质量的文献加以借鉴，切忌视野狭窄、主观片面、草率从事、结论武断。

（三）确定研究的关键问题

由于人的精力和时间是有限的，应根据学术力量、知识的水平、研究的基础和研究的具体条件从诸多需要解决的问题中，参照有关选题的原则，选择研究的关键和切实可行的重大研究问题立题研究。对此，要实事求是，要估计研究成功的把握度，以充满信心地完成研究任务，达到预期的目的。这里切记的是不要贪多求全，企图通过一个研究课题来解决许多个问题，实践证明这样的想法和做法，都是脱离实际的。

（四）撰写研究计划书

立题后，应按一定规范格式和内容，撰写科研计划书。详见本教材第十六章。

四、临床研究立题的评价标准

对于临床研究立题的评价，有以下参考标准供研究者考虑：

（一）是否为国家或地区性危害人民健康的重大疾病

这里要考虑社会的需求性、注重国家或地区性疾病负担的现况，要重视对危害人民健康重大的疾病研究；要考虑考虑社会的公正性，即临床研究能为大多数人服务。当然，还要结合研究者和研究机构的实际水平与研究的能力，实事求是地为国家或地区有效防治重大疾病、研究创新方面作贡献。

（二）研究的重点是否明确

一个科研课题应有明确的研究重点，即重点要解决的关键问题。例如：对 SARS 特异性早期诊断研究，重点究竟是什么？是分离病毒确诊或是应用特异性抗体检查早期患者的病毒抗原？还是检验 SARS 患者血中的抗体？方法是什么？是否准确可信？预期价值如何？这些问题应明确回答，方可论证研究的重点。

一个研究课题，若涉及的问题面宽，想通过一个研究解决多方面的问题，必然缺乏重点，其结果也不会达到预期的目的。

（三）是否掌握了本研究涉及的最新科技信息和发展动态

立题研究与确立研究的重点，务必要通过检索文献资料，查阅与掌握相关的最新科技信息，这是避免无效重复及使得自己掌握最新知识，力求创新的最为重要的环节。应力争完善并对新知识做到严格评价，取其精华、弃其糟粕，掌握可靠的最新最佳证据方能在探索中继往开来，使立题研究更具时效性、科学性。

（四）是否具有创新性

一个好的研究课题，务必要有自主性的创新性，重复他人的“研究”是缺乏生命力的，如像有的研究人员从国外引进一点技术或试剂，照人家的方法在国内找一些患者采集一些标本，做一下人家做过的试验，获得某种结果，包括分子生物学的“研究结果”进行报道，尽管在国内算首次，但算不上“自主创新”，因为毕竟是重复人家的，创新应该是自己真正的创造。此外，亦应

避免盲目地与国际"接轨"和脱离国情的"追新"！所以在创新上应认真鉴别，尽管创新难，但在研究工作方面又确属重要。

（五）是否具有可行性

对课题承担与执行者的学术技术力量、工作基础、设备条件、研究设计的方案和路线、干预措施、研究对象的来源和数量上的保障以及经费的预算，是否可行，要仔细评价。

（六）预期的成本 - 效果

这里要考核研究课题预期可能产生的效果。例如提高诊断的水平、提高疗效水平，降低病死率、病残率、伤残调整寿命年等的指标以及为此而投入的成本，计算其成本 - 效果和作相应的卫生经济分析，这样才可能看出一个研究课题所产生的社会和经济的效益。

（七）医学伦理

临床科研务必要保障患者的安全及权益，尊重患者的人权和人格为最高伦理标准，而课题的立题应通过伦理学组织的审查批准方可实施。对此，任何临床试验都应遵照世界医学会所制定的赫尔辛基宣言的精神。

（刘晓清　王吉耀）

笔记

第三章　临床研究设计的基本要求

临床研究是以病人及相应群体为研究对象，研究和论证某个或某些研究因素对疾病病因、诊断、治疗、预后等方面产生的效应或影响。与理化实验及动物实验不同，人体作为最高形式的生命体，除生物学属性存在个体差异外，在心理和社会功能属性方面也有不同；另外，就疾病本身而言，病程、病情、诊治手段等方面的差异，也进一步增加了临床研究的复杂性。因此，临床研究的设计非常重要，掌握临床研究设计的基本原则和基本要求，将有助于生产高质量的临床证据。

第一节　随机化原则

一、概　　述

随机化（randomization）是采用特殊手段，使总体或样本中每个个体发生某事件的概率均等。在临床研究中，随机化主要应用在抽样和分组两个环节，即在选取样本和将研究对象分组时，为防止来自研究者与被研究者两个方面主观因素的干扰，避免结果失真而采用随机化。因此，随机化可分为随机抽样与随机分组两种形式。

1. 随机抽样（random sampling）　由于临床研究的目标总体常常是无限总体，如罹患某疾病的所有患者或发生该疾病的潜在人群，直接进行总体研究的可行性很小；另一方面，即使是有限总体，由于人力、财力、时间等因素的限制，也不可能把所有对象都纳入研究，因此，只能从总体中抽取一定比例的样本、作为总体的代表进行研究。随机抽样是在抽样过程中，采用随机化方法，使总体中所有对象都有同等的机会被抽中进入研究样本。随机抽样的目的是保证样本的代表性，避免发生选择偏倚。但在很多临床研究中，研究对象来源于医院门诊及住院患者，数量有限，无法进行随机抽样，而只能连续纳入某段时间内的病例作为研究样本。

2. 随机分组（random allocating）　随机分组又叫随机分配，它是指在研究样本确定后，进一步采用随机的方法，将研究对象以同等的机会分配进入试验组（experimental group）或对照组（control group）中。通过随机分组，可以提高组间的均衡性，减少非研究因素的干扰。

二、随机化方法

随机化原则是针对每个独立观察单位（sampling unit）进行随机抽样或随机分配。因此，在实施随机化方法之前，应根据研究目的及相应的设计方案确定研究的基本单元，即研究的独立观察单位。临床研究通常是以单个个体（如某病患者）为独立观察单位。但是，有时因研究目的或研究因素的特殊性，不能以单个个体为独立观察单位，而需要采用某一特定组群（cluster），如一对夫妻、一个家庭、一个班组，甚至一个社区作为独立观察单位。如探讨生殖健康教育与育龄夫妻避孕效果的关系，就常以一对夫妻作为一个独立观察单位。另外一种情况是以个体的一部分器官或组织作为研究基本单元，如眼睛。

（一）简单随机法（simple randomization）

简单随机法主要指采用抛硬币、抽签、掷骰子、查随机数字表，甚至利用计算器、计算机产生的随机数字进行抽样或分组的方法。

笔记

1. 抛硬币、抽签法　该法主要用于小样本研究。例如欲在10个观察单位中抽取两个作为

研究对象，可制作10个签，其中8个为"无"，2个为"有"，抽中"有"者作为研究对象。在随机分组时，可事先规定硬币"正"、"反"面中的一面为试验组，另一面则为对照组，然后将研究对象依次按照硬币抛掷的"正"、"反"面，进行简单随机分组。但要注意在样本较小时，硬币抛掷出的"正"、"反"面频率不一定为50%，有时可能出现较大偏离，使得不同试验组间观察对象数也存在较大差异，需进一步调整。而当遇到大样本时，利用上述方法进行抽样或分组，操作较为繁杂、不太方便。

2. 随机数字表法　查阅随机数字表法是实现随机化的一种较为简单、实用的方法。数理统计学家基于概率论原理编制了随机数字表（附表），表中排列的数字无论以行、列或斜向顺序都呈随机状态。利用随机数字表，既可以进行随机抽样，也可以完成随机分组。

（1）随机抽样：例如，欲在某医院门诊随访的250名高血压患者中随机抽取20名作为一种降压新药的试验对象。先将250名高血压患者从1~250顺序编号，然后从随机数字表任一数字开始，按照行/列或斜向顺序产生20个00~99之间的随机数，每个数字乘以250取前3位数，重复的数弃去重抽，最后，以对应编号的患者组成参加本次试验的样本。但应注意，一旦确定行、列或斜向顺序后，中途不能任意更改方向。

（2）随机分组：例如，欲将符合试验要求的20名患者，随机分配为试验组（A），对照组（B）2组。首先将患者从1~20编号，然后从随机数字表的任何一行、任何一列开始，顺序产生20个00~99之间的随机数。按事先规定，对应的随机数为"奇数"者进入试验组，为"偶数"者进入对照组。如从附表"随机数字表"第9行，18列开始，按从左到右，从上到下顺序获取20个随机数字并依次列入表3-1。

表3-1　20例患者查随机数字表分配情况

患者编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
随机表数	52	77	65	15	33	59	05	28	22	87	26	07	47	86	96	98	29	06	10	58
归组	C	T	T	T	T	T	T	C	C	T	C	T	T	C	C	C	T	C	C	C

如果结果显示两组例数不等，则应该进一步遵循一定的随机方法，对两组的例数进行调整。

3. 计算机或计算器随机法　该方法是大样本研究中最常用的一种方法。大多数具有编程功能或数字计算功能的计算机软件都可以通过随机数发生器产生的随机数。计算机产生的随机数是取值在0~1之间（不包括0和1本身）均匀分布的随机变量。例如，欲产生5个0~999的随机数，一般需事先指定一个种子数（seed），相当于在随机数字表上指定行和列。可在SAS（6.12版本）中指定种子数为8888，产生5个随机数：0.947 32，0.148 50，0.638 43，0.535 16，0.203 71。因计算机产生的随机数是在0~1之间，要得到0~999的随机数，只需将每个数乘以1000并取整，据此，以上5个随机数转换为：947，148，638，535，959。同一软件用相同种子数所产生的随机数是一样的（统计学上称为重现性），而不同软件所得结果不一样。用函数计算器的"RAN"键，也可以产生取值在0~1之间的随机变量。每按一次"RAN"键，可产生一组随机数字。

得到随机数字序列后，可按照类似随机数字表法进行随机抽样或随机分组。

必须指出，交替将受试对象分到试验组和对照组的分配方法（TCTCTC……）不属于简单随机分配法。因为，当研究者预先知道下一位患者将分配到哪一组时，可能在主观上对患者的病程早晚、病情轻重等进行一定的取舍，导致研究对象选择上的偏倚。

（二）分层随机法（stratified randomization）

1. 分层随机抽样（stratified sampling）　该方法是先将研究对象按某一特征进行分组（层），然后在各层中采用简单随机的方法抽取研究对象组成样本。在进行分层时，往往是以对观察值变异影响较大的因素作为分层因素（stratifying factor）。例如，进行某地区糖尿病患病率调查时，

笔记

可以以城乡、年龄、性别等对糖尿病患病影响较大的因素作为分层因素。分层随机抽样可以保证各“层”都有一定研究对象进入样本，提高了样本的代表性。

2. 分层随机分组(stratified allocating) 分组时，也是按照研究对象按某一特征先进行分组(层)，然后再在各层中按简单随机的方法，分配试验对象与对照组对象，最后将各层试验对象合在一起作为试验组、将各层对照对象合在一起作为对照组。在分层随机分组中，主要以研究对象中某些可能产生混杂作用的特征作为分层因素，如研究对象的重要临床特征或预后因素(包括年龄、性别、病情、有无合并症等)。在临床研究中，可基于以下原则设置分层因素：①选择所研究疾病或其并发症的危险因素；②选择对所研究疾病预后有明显影响的因素；③遵循最小化原则，将分层因素控制到最低限度，否则分层过多，会造成组内研究对象过度分散。

例如，慢性心房纤颤复律后进行抗心律失常药物维持治疗的随机对照研究。考虑到慢性心房纤颤患者的预后与病因、心脏大小及心房纤颤病程长短有密切关系，因此，宜针对上述三个因素(①病因：瓣膜性心脏病与非瓣膜性心脏病；②心脏大小：心胸比例≥0.50 和 <0.50；③病程：心房纤颤病程≥6 个月及 <6 个月)进行分层随机分组(图 3-1)。

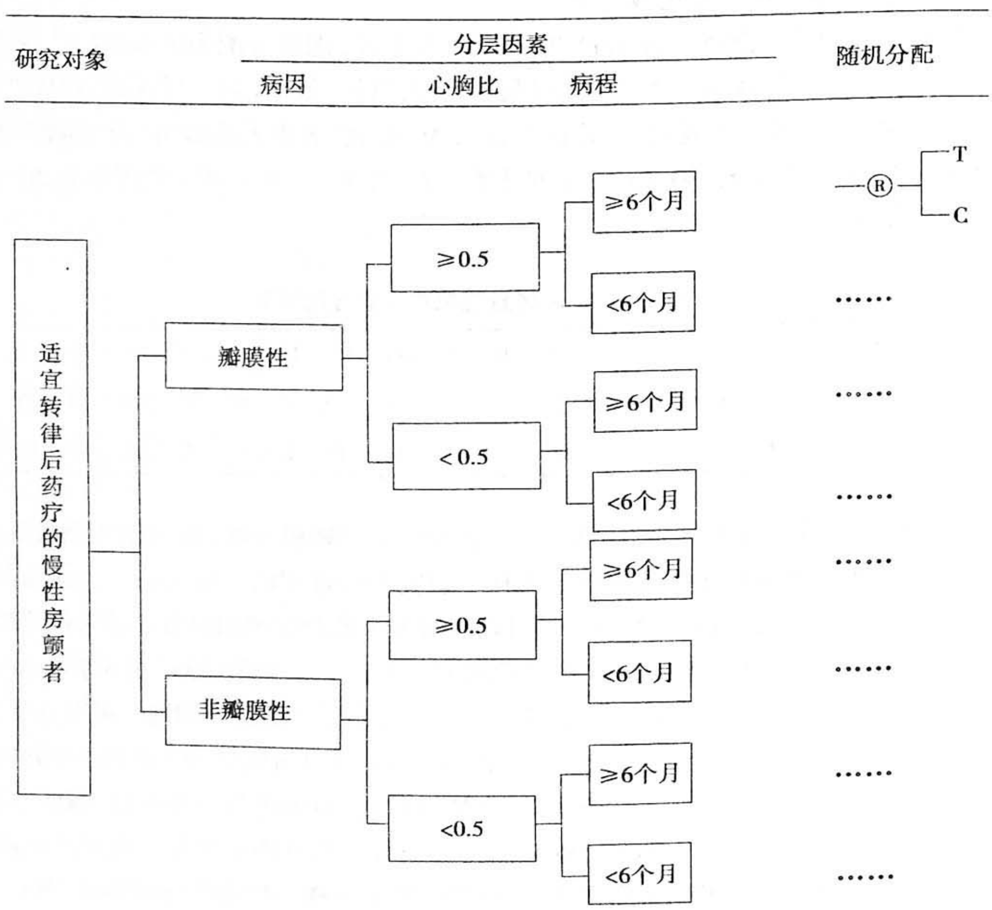


图 3-1 分层随机分组

经图 3-1 所示的分层随机分配过程，可以使病因、心脏大小、病程这三个对治疗效果和预后有较大影响的因素在试验组(T)与对照组(C)间保持平衡。

分层随机分组可以保证各“层”都有对象进入试验组或对照组，提高了试验组间的均衡性。但在大样本临床试验中，当样本量达数百甚至上千时，可不考虑分层随机分组。一方面，大样本经过简单随机分配后，不同试验组间的基线状况往往具有可比性；另一方面，需要有完整的研究对象名单后才能进行分层随机分组，但对大样本而言，可操作性不大。如果在资料分析时发现试验组之间均衡性不好，还可以通过分层分析进行统计学处理。注意，分层随机分组与分

层分析是两个不同的概念和方法。分层随机分组是在设计阶段，对一些可能对研究结果产生较大影响的非研究因素而采取的一种控制措施，其目的是增强组间的可比性。而分层分析则是在试验结束后，针对某些可能对研究结果产生影响的因素而采用的一种统计分析方法。当样本量较小、分层后有些层组可能出现对象数为“零”的情况时，应慎用分层分析。

（三）区组随机法（block randomization）

区组随机法主要用于随机分组。区组随机分组是先将研究对象分为不同区组，然后再对每一区组内的个体进行随机分配。在临床研究中，每一区组的研究对象数一般按组别的固定倍数来确定。如研究分为试验组、对照组两组，则区组例数可选 2、4、6、8 等，但区组例数越大，研究对象在分配时的排列组合越复杂。例如某研究分为试验组（T）、对照组（C）2 组，确定区组例数为 4。首先，研究对象按进入试验的先后顺序，每 4 个人一组，然后再对每一区组的 4 个研究对象分别根据随机数字表进行随机分组。区组随机分组的特点之一就是分组后各组人数相等。但当样本量偏小时，也可能出现 2 组例数不等的情况，必须进行适当调整。

区组随机分组有两方面优点：①保证各组人数相等；②便于逐渐累积临床病例。如上例，每积累 4 例研究对象即可进行分组并开始试验，不需要把所有样本全部收集齐后再进行分组研究。

（四）系统随机抽样法（systematic sampling）

系统随机抽样简称系统抽样或机械抽样（mechanical sampling）。即先将总体的观察单位按某一特征的顺序（如按入院先后顺序）编号，再根据抽样比例将其分为若干部分，先从第一部分随机抽取第一个观察单位，然后按一固定间隔在第二、第三……等各部分抽取观察单位组成样本。例如，欲从 2000 个观察单位中抽取 100 个组成样本，即抽样比例为 5%（抽样间隔为 1/20），可先从第 1~20（第一部分）之间随机抽出一个观察单位，如为 12 号，此后按每隔 20 抽取一个单位，即 32、52、72……至 1992 号组成样本。

系统抽样的优点是简单易行，适用于大样本的流行病学调查，样本的观察单位在总体中分布均匀，抽样代表性较好。缺点是如果总体各单元的排列顺序存在一定周期性，以这种方法进行抽样则可能出现较大的偏倚。例如，某工厂工作证编号中末位数代表车间，用该法按工作证顺序抽样则可能抽到的全部是同一车间的对象，使样本不能代表全厂的情况。

（五）整群随机法（cluster randomization）

整群随机法是以现成的群体（社区、街道、乡、村、医院、病房等）而不是个体为单位，进行抽样或分组。在整群随机抽样（简称整群抽样）中，抽到的群体中的所有观察单位，都将作为研究样本。例如，采用整群抽样方法，从某地区的 20 所小学中随机抽取 2 个学校，并对这 2 个小学的全部学生进行视力检查，以了解该地区小学生近视率。同样，在整群随机分组中，被分到试验组的群体中的每个观察单位，都作为试验对象，被分到对照组的群体中的每个观察单位，都作为对照对象。

采用整群随机法要求群间的变异越小越好，否则将影响样本的代表性或组间的可比性。一般情况下，用相同的样本含量，整群抽样的抽样误差最大，整群随机分组的组间可比性最小。但是，整群随机法具有节约人力、物力、方便、容易实施等优点，在实际工作中可行性较好，适用于大规模研究。

（六）多级抽样法（multi-stage sampling）

多级抽样是一种从大到小多个级别进行的抽样方法。首先从总体中随机抽取范围较大的单元，称为一级抽样单元（例如省、市），再从抽中的一级单元随机抽取范围较小的二级单元（如区、街道），若抽样到此为止称为二级抽样，若再继续往小范围抽样，则称为多级抽样。

在一些大规模研究（全国高血压抽样调查、全国糖尿病患病率调查等）项目中，多级抽样成为唯一实用的抽样方法。在具体实施的时候，多级抽样常常与上述各种基本抽样方法综合使用。

笔记

(七) 半随机化法(quasi-randomization)

不按随机数字产生序列号,而是根据被纳入研究对象的出生日期、住院号或门诊号等分为单、双数,分别将单数者分为1组,双数者分为另一组,这种方法称半随机化方法。

当进行大样本研究时,如社区人群的试验,利用半随机化分配受试者既简便、实用,同时也能较好地实现组间均衡。但在临床试验中,研究对象往往较少,半随机化很难达到组间均衡,因而该方法的使用受到限制。

三、随机化分配的优缺点

(一) 优点

1. 随机分配保证了研究组之间的均衡、可比。特别是采用分层随机分配,可使一些影响疾病进程的重要临床特征,以及一些已知可对研究结果产生影响的混杂因素在研究组间保持均衡,有效避免了这些因素对研究结果的干扰。

2. 随机分配可以防止选择偏倚和来自研究者、被研究者两方面主观因素的干扰,确保了研究结果的真实可靠。

3. 随机分配结合盲法一起使用,可以使研究结果更加客观,大大提高研究结果的准确性。

(二) 缺点

1. 在随机分配过程中,为了保证组间的均衡、可比,往往使研究对象局限于符合条件的患者,因而可能导致研究对象代表性的降低,使研究结果的推广范围受到局限。

2. 随机化分配往往需要累积一定数量的研究对象后才能实施,增加了临床研究的难度。

3. 随机分配可能违背研究对象的个人意愿,存在医德问题。

四、随机化过程中需注意的问题

在使用随机化方法时,应注意以下问题:

1. 随机化分组时,必须对随机化方案进行保密,该过程称为随机化分配的隐匿(concealment)。从某种意义上讲,随机化分配的隐匿比单纯产生随机分配序列本身更加重要。如果随机分配序列产生者与试验执行者(分配受试对象入组者)为同一人,试验执行者就可以预知下一位受试对象将被分入哪一组。在这种情况下,很难避免因研究人员主观因素造成的对受试对象的刻意选择,甚至改变分配方案,所带来的偏倚。因此,随机化分配方案的隐匿就是让分配受试对象入组的试验执行者事先不知道随机分配方案的具体内容,无法预知下一位入选的对象将进入哪一组。实现随机化分配隐匿的方法有中心电话随机系统、药房控制随机分配方案、利用编号或编码的药品容器、按顺序编码、密封入不透光的信封等,研究者可根据具体情况选择应用。

2. 为了保证实验的可靠性,研究中所用随机化方法、随机数等均应详细记录。如果是用计算机产生的随机数,需要报告产生随机数的程序、种子数等;如果是从随机数字表获得随机数,必须说明所用随机数字表、起始页、起始行、起始列及获取方法。特别是在新药的临床试验中,随机数必须具有重现性,产生随机数的参数及程序应与盲底一起封存。

第二节 对照的原则

一、概 述

所谓"对照(control)",指的是在临床研究过程中,设立条件类似、诊断一致的可供相互比较的组别。

笔记

"有比较才有鉴别",这是辩证唯物主义认识事物的基本法则。临床研究主要是探讨某个或某些研究因素对疾病病因、诊断、治疗、预后等方面产生的效应大小及其安全性,但由于临床事件的发生有着很大的不确定性,因此,需要通过对照比较才能得出可靠结论。

二、设置对照的意义

在临床研究中,除了干预措施(处理因素)的效应外,还有很多因素可能左右临床结局:

1. 不能预知的结局(unpredictable outcome) 由于个体的生物学变异和社会、心理因素的影响,导致临床上患相同疾病的个体,其临床特征(临床表现、病情轻重、预后等)存在较大差异。例如感冒、细菌性痢疾、病毒性肝炎等许多传染性疾病都有明显的自愈倾向,而在高血压、糖尿病、胃溃疡等慢性病的病程中,也存在病情的自行缓解或加重现象,即使是恶性肿瘤,其生存时间长短、并发症的发生等情况也各不相同。

2. 霍桑效应(Hawthorne effect) 指人们因成为研究中特别感兴趣和受关注的对象而产生的一种正向心理——生理效应,这种效应与他们接受的干预措施的特异性作用无关。霍桑效应一般是一种正向效应,即与干预措施预期效果同向的效应。但有时也可能因为厌恶某医生或不信任某医疗单位而产生负向效应。

3. 安慰剂效应(placebo effect) 某些研究对象,即使仅使用了安慰剂,也可能表现出病情好转等不该有的效果,其实质是某些研究对象由于医药依赖而产生的一种正向心理效应。

4. 向均数回归现象(regression to the mean) 指一些极端的临床症状和体征有向均数接近的倾向。如血压水平处于高限5%的人,即使不给任何处理,经过一段时间后,其血压值也可能有所降低(向均数靠近)。

5. 潜在未知因素的影响 由于知识水平和认识的局限性,很可能还存在一些影响干预效应的因素,但目前尚未被人们所认识。

因此,在临床研究中设立合理的对照非常必要,其主要目的是消除非处理因素的干扰,鉴别试验性与非试验性效应,减少或消除试验误差。

三、对照的种类

(一)按照研究的设计方案分类

1. 同期随机对照(concurrent randomized control) 按严格规定的随机化方法将研究对象同期分配到试验组和对照组,以此方案设置的对照类型称为同期随机对照。

同期随机对照的优点主要有三方面:①随机化分组,可以较好地保证试验各组之间的均衡可比,有效避免了潜在未知因素对试验结果的影响;②设置同期对照,可以同时对各试验组进行观察,有效避免了因试验先后顺序对结果的影响,使研究结果更有说服力;③由于多数统计方法都是建立在随机样本的基础之上,采用本设计类型更有利于资料的统计分析,不需要对数据进行校正。同期随机对照的缺点是:①需要有一半对象充当对照,因此所需样本量较大;②必须收集到样本含量所需的全部对象后才能进行试验,增加了研究的难度;③在有些情况下可能涉及医德问题。

2. 自身对照(self control) 将一组受试对象分为前后两个阶段,分别施加不同的干预措施,然后比较两个阶段的两种处理效应的差异。一般在前一阶段结束后应有一段时间间隔,称为洗脱期。设置洗脱期的目的是为了避免前一阶段的处理效应对后一阶段产生影响。

自身对照主要适用于慢性反复发作性疾病的治疗性研究,这一类疾病病程长,病情变化不大,允许进行分阶段处理与观察。采用自身对照可以消除个体差异,减少一半样本量,并且保证每个受试对象都接受有效治疗的处理;缺点是难以保证两个阶段的病情完全一致,可能存在处理先后对结果的影响。

笔记

3. 交叉对照(cross-over design control) 将两组受试对象分两个阶段进行试验。第一阶段一组试验用A措施,间隔一段洗脱期后再试验B措施;另一组第一阶段试验B措施,间隔一段洗脱期后再试验A措施,然后对比A、B两种措施的效果。

交叉对照同样可以保证每个受试对象都接受有效治疗的处理,消除个体差异,节约样本量,提高统计效率,同时可以避免因试验先后顺序对结果的影响;但是,该方法的应用也主要限于慢性复发性疾病。

4. 配对对照(matching control) 将可能对研究结果产生影响的混杂因素(如年龄、性别、病情等)作为配比条件,为每一个试验对象选配一个以上的对照,通常采用1∶1或1∶2配对。

配对对照可以最大限度地保证组间均衡性,避免已知混杂因素对结果的干扰。

5. 非随机同期对照(non-randomized concurrent control) 有同期对照,但试验组与对照组未严格按随机化原则进行分组。例如在多中心临床试验中,将一个医院的对象全部作为试验组,而另一个医院的对象全部作为对照组进行研究。

非随机同期对照的设置简便易行,易为医生和患者接受;缺点是由于非随机分配,可能因选择偏倚导致两组基线情况不一致,可比性较差。

6. 历史对照(historical control) 试验仅设一组接受新干预措施,并与过去采用的旧干预措施进行比较。因此,历史对照属于非随机、非同期对照。

历史性对照的资料可以通过查阅医学文献和医院病历记录获得。优点是节省时间、经费,不存在医德问题;但通常既往的资料中的研究对象、条件、环境等都很难与本次试验保持一致,可比性差,一般不宜采用。在特殊情况下,如对一些预后极差的疾病,采用历史对照还是有一定说服力的。如恶性滋养层细胞瘤在20世纪50年代单用手术治疗,预后极差。随着化疗和放疗的进展,联合化疗的有效率达70%~90%,对这两个历史阶段中,同一疾病的不同治疗效果进行历史对照,证明现在的治疗效果明显提高,预后显著改善,为人们所公认。

(二)按照对照组的处理措施分类

在以上各种对照设置方案中,试验组的处理一般是施加某种需要评价的干预措施(处理因素),而对照组也需要采用相应的对照措施(对照因素)。目前使用较多的对照措施有有效对照(阳性对照)、空白对照、安慰剂对照等。

1. 有效对照 即以目前临床公认的有效处理方法(如治疗某病常规、有效的治疗方法)施加给对照,然后与试验组处理措施(新治疗方法)的效果相比较。这类试验通常采用随机双盲设计,受试者随机地分配至试验组和有效对照组。是临床治疗性研究中最常用的对照方法。

在有效对照中,施加给对照组的处理措施效果稳定,试验期间能保证对照组成员受到合理治疗,较少引起伦理学及医德方面的问题。

2. 空白对照(blank control) 对照组在试验期间不给任何处理,仅对他们进行观察、记录结果,并将其与试验组的效应进行比较。

在空白对照中,对照组成员在试验阶段未得到任何治疗,可能造成不良后果,应用时要特别谨慎。空白对照不能用于急、重或有较重器质性病变的病人,仅用于病情轻、稳定,即使不给任何治疗也不会导致病情恶化的疾病(如近视),否则将产生伦理和医德方面的问题。

3. 安慰剂对照(placebo control) 安慰剂是不具有真正治疗或致病效应的制剂。将安慰剂与试验措施进行比较称为安慰剂对照。常用的安慰剂为没有药理活性,但其色、形、味均与试验的新药完全相同的制剂,如淀粉、乳糖、维生素等。若注射给药,多采用生理盐水作为安慰剂。

虽然安慰剂对照本质上也是一种空白对照,但其效应不同。因为,使用安慰剂后可产生安慰剂效应。采用安慰剂对照的主要目的是为消除主观因素的影响,便于盲法的实施。安慰剂对照与空白对照一样,主要用于病情轻、稳定,或目前没有有效治疗方法的疾病。同时,也应注意伦理学和医德方面的问题。

笔记

第三节　盲 法 原 则

一、概　　述

盲法试验（blind trial）是指临床研究过程中，指标的观测、数据的收集和结论的形成等均在不知晓研究对象所在的组别以及不知晓所接受的是何种措施的前提下进行。此外，盲法还可用于统计分析之中。

临床研究要求对每一个研究对象的诊断、疗效或预后等作出准确、可靠的测量与评价，否则将出现测量偏倚（measurement bias），导致研究结果失真。在临床试验中，测量偏倚既可来自研究人员方面，也可产生于研究对象方面。一旦研究对象知晓自己的分组情况，可预知疗效或预后，产生霍桑效应，甚至会对治疗丧失信心而中途退出。另一方面，若研究人员知晓受试对象的分组情况，则可能因希望研究得出阳性结论而有意或无意识地暗示受试对象，诱导治疗组对象多报告有利结果，对照组对象多报告不利结果。进行盲法试验的目的就是为了克服可能来自研究者或受试者的主观因素所导致的偏倚。

二、盲法的分类

盲法试验的基本原则就是让受试对象和（或）研究人员无法知晓哪些人接受的是处理措施，哪些人接受的是对照措施。在一项临床研究中，往往涉及研究对象、干预措施执行者、结果测量者、统计分析者和论文撰写者等，根据"盲"的对象不同，一般可将盲法分为单盲、双盲、三盲等3种类型。

（一）单盲

1. 定义　对于研究对象的分组及所施加的处理因素（如选用药物）情况，只有研究者知道，而受试对象不知晓，叫做单盲（single-blind）。所以，"单盲"主要是盲研究对象。另外，对检验人员、病理学检查人员等隐瞒患者的诊断与治疗情况，便于他们客观报告试验结果，也属于单盲。

2. 优点

（1）方法简便易行。

（2）干预措施执行者和结果测量者知道受试对象的分组情况，便于更好地观察、掌握病情。一旦病情发生变化，可以及时、恰当调整处理方案或采取其他干预措施，使受试对象免受伤害。

（3）减少了因受试对象主观因素对研究结果的影响。

3. 缺点　单盲不能消除研究方主观因素造成的影响。主管医生可能通过不同途径去影响患者的疗效，比如，医生对接受新疗法的患者观察特别仔细，护士对新疗法组患者更加关心和热情，这些都可能影响或暗示受试对象产生不同的反应。

（二）双盲

1. 定义　受试对象和试验执行者（干预措施执行者及结果测量者）双方均不知晓分组情况，也不知道受试者接受的是何种干预措施，称为双盲（double-blind）。

2. 实施方法及注意事项

（1）双盲试验往往与安慰剂和随机化分配的隐匿同时配合实施。随机化分配的隐匿是双盲试验的基础，没有随机化隐匿过程，则无法保证受试对象与试验执行者双盲，而设计科学、合理、逼真的安慰剂则是实现双盲所必备的条件。

（2）双盲试验要有严格的管理制度和方法，组织严密，操作规范。在双盲试验过程中，需要有"局外"的管理、监督者，他们不直接参加临床研究的观测和数据收集，仅参与研究设计、药物编码的控制与保密、资料的保管、分析等，而受试对象和研究的执行者只知道每个试验对象的

笔记

药物编号，如A和B。待试验结束并完成数据分析后，才由密码控制者当着主要研究人员的面，公布隐匿的随机化方案并宣布A药和B药究竟是试验制剂还是对照制剂。该过程称为“揭盲”。

（3）双盲试验要求各组的药物在外观的形状、大小、颜色，给药的途径、方法、次数上保持一致，若有疏漏，则可能导致泄密，失去“盲法”的作用。药物临床试验时，若试验药与对照药的剂型不一样，可采取“双盲双摸拟”方法：试验药＋与对照药剂型一样的安慰剂；对照药＋与试验药剂型一样的安慰剂，并分别编上密码。若制剂的效应维持时间不同，也可以采用“双盲双摸拟”方案进行。例如，用双盲法评价两种抗高血压药的效果。一组用缓释的抗高血压药，每日一次；另一组用β受体阻滞剂，每日两次。在执行时可采用两组患者每天早、晚各给药一次。药名标出A和B，两组患者均为早上服A药一片，晚上服B药一片，而两组患者所用A药和B药是不同的。具体做法如下：每日服药一次组：A药＝慢释放的抗高血压药；B药＝安慰剂。每日服药两次组：A药=β受体阻滞剂；B药=β受体阻滞剂。

（4）在盲法试验过程中，一旦某一受试对象的病情突然发生变化，如出现严重的副作用、治疗无效或病情加重等，必须立即停止试验，予以“破盲”，并同时使用必要的治疗措施。因此，在盲法试验进行之前，需事先制定出明确的停药或更换指征，以免给受试者带来不良后果。同时，“破盲”应局限在较小范围内进行，以尽量减少因“破盲”带来的影响。

3. 优缺点

（1）优点：是临床试验最常用的一种盲法形式，可以有效避免受试对象和试验执行者主观因素对试验结果的影响。尤其对干预措施的评价指标是以主观指标（如恶心、呕吐、头痛、眩晕、呼吸困难等）为主时，采用双盲试验能够有效减少测量偏倚的发生。

（2）缺点：①在管理上缺乏灵活性；②有特殊副作用的药物容易被破盲；③双盲试验不适用于危重患者。

（三）三盲

1. 定义 三盲（triple blind）是指，受试对象、试验执行者和资料分析与报告者三方均不知晓受试者接受的是何种干预措施，全部采用编号密封，它可避免双盲法在资料分析阶段的测量偏倚。

2. 优缺点

（1）可以使偏倚减到最小的程度，使评价结果更趋真实。

（2）设计复杂，执行难度较大，常因医德、沾染、补偿、失密等问题而难以坚持。

（3）当受试对象的病情突然发生变化时，难以及时掌握治疗情况，造成处理延误。

所以，尽管三盲试验是减少偏倚最有效的方法，但在实际工作中使用并不普遍。在许多临床研究中，由于临床医师既是试验设计者与观察者，也是资料分析和结果评价者，很难真正做到三盲。

三、非盲法评定

（一）定义

非盲法评定，又称为开放试验（open trial），即受试对象和研究者均知晓试验组和对照组的分组情况，以及所给予的干预措施。

并非所有临床研究都适用盲法，例如，比较手术疗法和保守疗法的治疗效果，探讨功能训练或针灸疗法的疗效，评定生活习惯（吸烟、运动、饮食）对疾病的影响等。在这种情况下，盲法试验难以实施，只能采用非盲法评定；另外，若主要结局指标是明确的硬指标（如存活或死亡），很少受主观因素影响，采用非盲法评价也可以获得比较满意的结果。

（二）适用范围

笔记

1. 在一次研究中，需要对多种因素进行分析。如不仅希望观察药物的疗效，同时也要评价

影响疗效或发病的有关因素。

2. 研究者希望更安全、更周到地作出决策（例如患者是否需要继续治疗，药物是否需要增减，是否需要配合其他药物等），使医疗决策更灵活。

（三）优缺点

1. 易于设计和实施，研究者了解分组，便于对受试对象及时作出处理。

2. 最主要的缺点是容易产生测量偏倚。

第四节　临床研究的伦理学基础

一、概　　述

科学技术方法在21世纪取得了迅猛发展，并越来越多地运用于医学研究各领域，这对增进人类健康、提高生活质量、预防和治疗疾病等均发挥了巨大的作用。同时，以人体为研究对象的临床研究也极易引发一系列伦理学问题。从Tuskegee梅毒研究、Willowbrook肝炎研究等案例来看，违背医学伦理的科学研究，无论研究结果如何，终将会受到全人类的谴责。

医学伦理的基础是由《希波克拉底誓言》发展而来的《日内瓦宣言》，该宣言对一名医生的医疗行为提出了伦理学准则并在1948年世界医师协会大会上正式通过，而在1975年世界卫生大会上提出的《赫尔辛基宣言》则针对医学研究中的伦理学问题做出了规范。20世纪90年代后，我国国家食品药品监督管理局结合本国实际情况，发布的《药品临床实验管理规范（GCP）》对药品研究中临床实验阶段所涉及的伦理问题也做出了明确规范。

临床研究以人体（病人）为受试对象，在研究过程中不可避免地会涉及伦理学问题。但每项医学研究都必须在获取科学结论与维护受试者利益间寻找一个最佳平衡点，使医学科学真正造福于人类。

二、临床研究立项中的伦理基础

临床研究中无论是新的预防、诊断和治疗方法效果研究，还是病因与病理生理研究都必须遵循“受益大于风险”原则，即有利于增进人类健康，尽可能避免各种潜在风险的基本原则。

为了切实达到上述目标，在研究立项前，应做大量的文献复习，并进行循证分析与决策。只有在大量证据证明研究的必要性及预期获益超过其所带来的风险和损害时，研究才有了立项的基础。

研究设计完成后，必须报伦理委员会审批。报送材料中应对所有涉及人体环节的研究过程、实施方法、潜在风险及控制措施等方案真实、完整呈现，以便委员会专家做出客观判定。

同时，临床研究者必须由合格的科研人员承担，并至少在1名以上具有丰富临床经验的医生监督下进行，以确保病人的安全。

三、临床研究中的伦理学原则

（一）尊重（respect）原则

尊重原则的主要内涵在于，医学研究应充分尊重受试者的人格及其作为一个自主个体所应享有的权利。尊重原则具体体现在以下几方面：

1. 知情（informed）　研究者应向受试对象阐明研究的目的、方法、实施过程、经费来源及研究可能带来的利弊风险，让受试者有充分的时间了解加入研究项目对其自身的意义和价值所在，以便做出理性的决定。

2. 自愿（voluntary）　对于有自主行为能力的受试者，必须在其完全自愿的情况下签署《知

笔记

情同意书》(*Informed Consent Form*),并告知他们有拒绝参加研究或无条件随时收回同意书的权利。对于自主性受限的对象(儿童、昏迷者、痴呆者等),研究者必须按法律规定从合法代理人处获取知情同意。同时,在临床研究中为避免受试对象(病人)因迫于研究者(医生)压力而被迫同意参与试验的情况发生,应由一位了解情况但不参与研究的独立第三方来征询对象的知情同意。

3. 保密(confidential) 研究者应对受试对象的个人信息、健康状况、行为习惯等资料严格保密,未经本人同意,不得将受试对象的个人信息公开,应充分尊重其隐私权。

(二) 受益(beneficence)原则

受益原则主要体现在两个方面:

1. 在研究过程中,应尽量保证受试者受益最大化、风险最小化。如尽量避免侵入性操作,抽血量降至最低。医务人员应密切观察受试对象的病情变化,一旦发现潜在风险大于可能的受益或已经得出预期结果,应立即终止试验。

2. 研究得出确切结论后,应确保每个参加试验的病人能够利用课题所证实的最好的预防、诊断和治疗方法。如在疗效研究中一旦证实试验组所用的新药疗效优于常规药物,则在试验结束后应该让对照组患者也使用新药。

(三) 公正(justice)原则

公正原则要求研究者在确定研究对象时,应该遵循科学原理,设立合理的入选标准、排除标准和剔除标准,不应该只是针对弱势群体(老幼、残障、囚犯、有色人种等)。其次,在分配受试对象到不同试验组时,应不分亲疏,一视同仁。分配前应告知对象可能被分配到试验的不同组别,并说明分组原则与方法。

第五节 临床研究中的偏倚及质量控制

一、概 述

除少数直接以目标总体为对象进行的研究(某有限总体疾病或健康情况普查、特殊病例报告等)外,绝大多数临床研究为抽样研究,即用样本结果来推断总体情况。在生物多样性背景下,样本测量值与总体真实值是不等的,这种差异在统计学上称误差(errors)。

医学研究中的误差来自两个方面:一是因抽样本身导致的误差,即统计学上的抽样误差(sampling errors)。抽样误差与抽样研究是相伴相随、不可避免的,该类误差因没有方向性且呈随机变化,又称为随机误差(random error),可以通过扩大样本量或提高随机抽样质量来尽量使抽样误差降至最低,统计学上常用标准误描述其大小(详见统计学相关章节),以帮助正确判断抽样误差对结果的影响。

导致误差的另一方面原因是偏倚(bias)。偏倚指抽样误差以外的,因研究设计、测量、资料分析及结果解释等临床研究各个环节中所发生的系列错误,导致研究结果系统性偏离真实值。偏倚导致的误差又称为系统误差(systematic error),有方向性,即某种错误导致样本均数 / 率系统性高于或低于总体参数,也可能是系统性夸大或减弱研究因素与结果间的关联强度,从而影响结果的真实性。偏倚所致的误差是不能通过扩大样本量来控制的。由于偏倚是研究中的各种错误所造成的,按其产生的原因可将偏倚分为选择偏倚、信息偏倚、混杂偏倚三类。

二、选 择 偏 倚

(一) 选择偏倚的概念

选择偏倚(selection bias)是指在选取研究对象时,由于选取的方法不当,使入选的研究对象与未入选者在某些特征上存在着差异而造成的一种系统误差。

笔记

选择偏倚主要发生在研究的设计阶段，当确定研究样本或选择比较组时，未严格遵循随机原则，或设置的纳入 / 排除标准不恰当。如探讨高血压与血脂异常的关系时，研究对象全部来自心内科住院患者，或者采用自身前后对照设计评价某减肥药疗效时，研究对象全部为自愿者，以上两种情况均可能夸大研究因素与研究结果间的关联强度、产生正偏。

另外，在信息收集过程中也可产生选择偏倚，如无应答、失访例数较多而又未采取有效的处理措施，若仅以资料完整对象进行分析，也可造成结果高估。

（二）常见来源举例

1. 入院率偏倚（admission rate bias）　入院率偏倚又称伯克森偏倚（Berkson's bias），指当研究对象全部来自医院的住院或门诊病例时，可能因患有某种疾病或具有某种特征者入院机会较目标人群不同，而导致的一种偏倚。例如，研究呼吸道疾病与骨及运动器官疾病的相关性，由于同时具有呼吸道疾病及骨骼与运动器官疾病的患者入院或就诊机会较高，因此在医院选择研究对象时，会夸大两者之间的关联度，见表 3-2。

表 3-2　呼吸道疾病及骨骼与运动器官疾病相关性研究

呼吸道疾病	骨骼与运动器官疾病					
	一般人群			住院 6 个月以上患者		
	有	无	合计	有	无	合计
有	17	207	224	5	15	20
无	184	2376	2560	18	219	237
合计	201	2583	2784	23	234	257

$OR_{一般人群}=1.06$，$OR_{医院患者}=4.06$，可见，在一般人群中呼吸道疾病与骨骼肌运动器官疾病间相关性很弱，而以住院 6 个月以上者为研究对象时，两者间显示有很强的联系。

2. 现患 - 新病例偏倚（prevalence-incidence bias）　现患 - 新病例偏倚又称奈曼偏倚（Neyman bias），发生在研究对象全部或大部分为现患病例（老病例和新病例）时。现患病例与新病例相比，在病程、病情、病型等方面可能存在差异，特别是当中可能有较多"生物学上的强者"时，可出现病例选择偏倚。

3. 检出症候偏倚（detection signal bias）　指在病因研究中，某种疾病的一些早期症状可能与研究的暴露因素相关，因此，出现这些早期症状的人会及早就诊，并接受检查而导致该人群有较高的疾病检出率，以致得出暴露因素与某病相关的错误结论，或夸大两者之间的关联度。例如子宫内膜癌的病因研究，若研究对象来自医院门诊时，得出服用雌激素是主要病因的结论，但在普通人群中研究时，两者关联明显减弱。进一步研究发现，服用雌激素可引起阴道不规则出血，这部分人会及早地到医院就诊，因此，大大提高发现其患有子宫内膜癌的机会。

4. 排除偏倚（exclusive bias）　指研究中"无应答"、失访或"不依从"者较多，若直接剔除该部分人资料，仅用余下资料进行分析而带来的一种偏倚。这是因为被排除对象与余下的对象可能存在明显差异。如在疗效研究中，疗效差或不良反应较大者更容易失访而被排除在分析之外。

5. 易感性偏倚（susceptibility bias）　当研究对象过于集中在某一类特定人群，如以油漆生产车间工人为对象，研究油漆与皮肤过敏间的关系，可能会发现这一组人群过敏性皮炎的发生率并不比其他人群高。这里面应考虑纳入对象的特殊性，如因自然淘汰，能够留在该车间工作者可能多是对油漆刺激不易感者。

（三）选择偏倚的防止

在不同研究中因目的不同，产生选择偏倚的原因也不相同。防止选择偏倚主要从以下几方面入手：

1. 当研究目的确定后，需广泛查阅文献，反复讨论可能出现选择偏倚的各种情况，从源头

笔记

上加以避免；

2. 尽量在一般人群中选择研究对象，或设计多组对象以控制选择偏倚的发生；

3. 纳入对象或分组时，尽量做到随机化；

4. 尽量减少“无应答”、失访或“不依从”情况，一般要求控制在15%以内。若确有较多“无应答”、失访或“不依从”者，需从这类对象中随机抽取部分个体，千方百计收集到他们的资料，并将其与应答完整或随访完整者的资料进行比较，若两部分对象在影响研究结果的因素上无显著差异，则可用资料完整者资料进行分析，反之则表明存在偏倚的可能性较大。

三、信息偏倚

（一）信息偏倚的概念

信息偏倚（information bias）又叫测量偏倚（measurement bias）或观察偏倚（observational bias），指在收集对象暴露特征、随访及结局信息时错误地把一类信息归为另一类信息所造成的系统误差。

信息偏倚主要发生在研究的实施阶段，其产生的原因主要有三个方面：①来自研究对象。研究对象可能由于主观或客观的原因有意或无意提供不真实信息，如孕妇有意隐瞒饮酒史，慢性病患者忘记曾经服用某种药物等。②来自研究者。研究者也有可能因主观原因有意或无意地选择性收集与接收某些信息，造成信息偏倚。如研究者一般都希望得到阳性研究结果，对暴露者的暴露相关信息的挖掘和收集比对非暴露者更加深入、全面，从而造成夸大暴露与结局间联系的误差。③来自研究中使用的测量仪器或工具。如果研究使用的测量仪器或工具本身存在测量误差，则必然导致研究结果的系统误差。

（二）常见来源举例

1. **回忆偏倚**（recall bias） 在进行问卷调查时，被调查者可能因记忆不准确，对以往的疾病史、用药史或家族史等情况提供错误信息而导致偏倚。回忆偏倚较多无倾向性，如回忆吸烟情况，可能偏多，也可能偏少，其结果往往会减弱研究因素与结局间的关联强度，产生负偏。

2. **报告偏倚**（reporting bias） 指研究对象可能出于各种原因故意隐瞒或谎报某些信息而导致的偏倚。如一些心血管病患者故意隐瞒吸烟、饮酒史等，造成某药物疗效评价失真。

3. **暴露怀疑偏倚**（exposure suspicion bias） 暴露怀疑偏倚主要来自研究者。如研究某因素与一种疾病的关系时，研究者从主观上认为两者有联系并希望出现阳性结果，因此在收集基本情况或随访信息时，会下意识的过度挖掘病例组的暴露信息而忽略对照组的暴露信息。暴露怀疑偏倚往往会夸大研究因素与研究结果间的关联度，造成正偏。

4. **诊断怀疑偏倚**（diagnostic suspicion bias） 在病因研究或预后研究中，被研究者因被确定有某种暴露或患有某种疾病，在主观因素影响下，会有意无意地强调或判断自己已有相应症状或结果事件发生，从而导致夸大因果联系的错误结论。

5. **检测偏倚**（detection bias） 检测偏倚指因检查、测量仪器误差或操作不当而导致的系统误差。

（三）信息偏倚的防止

根据信息偏倚产生的原因，防止信息偏倚主要从以下几方面进行：

1. 研究者对拟进行的研究要事先制定明细的资料收集方法。如设计统一而明确的调查表，对调查内容或测量指标规定明确客观的标准，避免模棱两可的提问，尽量采用定量指标收集信息；

2. 研究正式实施前，对调查员进行精心挑选和严格培训，让调查员充分了解每个问题的设置目的、内涵，掌握收集信息的要领和技巧；

3. 尽量采用“双盲法”收集信息，避免因研究者或研究对象的主观因素带来暴露怀疑偏倚、

笔记

诊断怀疑偏倚等，影响信息真实性；

4. 研究实施前，对所使用的仪器设备进行校正；

5. 设计严格的质量控制手段，以及恰当实用的质量监测、控制措施(如调查员签名、复查等)。

四、混杂偏倚

(一)混杂偏倚的概念

混杂偏倚(confounding bias)也可简称为混杂(confounding)，是指由于有一个或多个潜在外部因素的存在，夸大或掩盖了研究因素与结局之间的真实联系，出现错误估计结果。这些潜在的外部因素(extraneous factor)称为混杂因素(confounding factor)。

混杂因素是导致混杂偏倚的关键，它具备三个基本特点：①混杂因素本身就是一个独立的、与研究结局有关联的因素；②混杂因素与研究因素之间也必须相关；③它不是研究因素与研究结局因果链上的一个中间环节。以病因研究为例，当探讨某暴露因素 E(exposure factor)与疾病 D(disease)的相关性时，另有 1 个因素 F(factor)，它既是疾病 D 的一个独立的危险因素，同时又与暴露因素 E 呈正相关，F 因素的存在意味着暴露组必然有较高的疾病 D 发病率。因此，无论 F 与 D 之间是否真正有因果联系，都会得出暴露组发病危险性高于非暴露组的结论。

例如，某队列研究探讨服用维生素 E 与罹患心肌梗死间的关系，通过随访得到表 3-3 所示数据。

表 3-3 心肌梗死与服用维生素 E 相关性研究

服用维生素 E	心肌梗死		合计
	发生	未发生	
服用	400	600	1000
未服用	600	400	1000
合计	1000	1000	2000

统计显示，$RR = 0.4/0.6 = 0.67$，显示服用维生素 E 者发生心肌梗死的危险性较小，在有统计学意义的条件下，可以得出服用维生素 E 对心肌梗死有保护作用的结论。但是这一结果可能受到“吸烟”这个外部因素的干扰。那么，在本研究中吸烟这个因素是否会导致混杂偏倚呢？可通过以下分析来判断：

首先，分析心肌梗死发生是否与吸烟有关。通过随访记录上的吸烟情况记录发现，2000 人中，吸烟者 1150 人，心肌梗死发病率约为 71%，不吸烟者 850 人，心肌梗死发病率约为 21%，进一步结合服用维生素 E 的情况，得到表 3-4 所示数据。

表 3-4 按吸烟与否分层分析心肌梗死与服用维生素 E 相关性

吸烟情况	服用维生素 E 组			不服用维生素 E 组		
	心肌梗死			心肌梗死		
	发生	未发生	合计	发生	未发生	合计
吸烟	240	30	270	580	300	880
不吸烟	160	570	730	20	100	120
合计	400	600	1000	600	400	1000

通过统计分析得到：$RR_{服用\ VitE} = 3.64$，$RR_{不服用\ VitE} = 3.88$，以上结果说明，无论是否服用维生素 E，吸烟与心肌梗死都有很强的联系，是一个独立的危险因素。

进一步分析吸烟与服用维生素 E 的关系，得到表 3-5 所示数据。

笔记

表 3-5 吸烟与服用维生素相关性分析

吸烟情况	服用维生素 E				合计
	服用	(%)	未服用	(%)	
吸烟	270	(27)	880	(88)	1150
不吸烟	730	(73)	120	(12)	850
合计	1000	(100)	1000	(100)	2000

计算得：$OR=0.05$，说明吸烟与服用维生素 E 这两种行为间存在明显的负相关，即服用维生素 E 者吸烟较少。

从表 3-4 和表 3-5 可知，在该研究中，吸烟这个因素满足了作为混杂因素的基本条件，可能引起混杂偏倚。

因此，判断某可疑混杂因素是否引起混杂偏倚，可通过比较含该因素时研究因素与结局间的粗(crude)效应值(如 *cRR* 或 *cOR*)与排除该因素后的调整(adjusted)效应值(如 *aRR* 与 *aOR*)来判断。若数值差别过大，说明存在混杂偏倚，以调整效应量作为研究因素的纯效应估计值。

(二) 混杂偏倚的控制

混杂偏倚是在进行组间比较时，因混杂因素在组间分布不均而造成的。如研究饮酒与肺癌的关系时，由于既饮酒又吸烟者居多，吸烟作为混杂因素，可导致肺癌与饮酒有关的错误结论或夸大饮酒与肺癌间的真实联系。因此，控制混杂偏倚主要从研究设计与资料分析两方面入手：

1. 设计阶段

(1)分组时必须遵循随机化原则。通过随机分组，可以相对保证混杂因素在组间分布均衡，从而起到控制混杂偏倚的作用。对于已知混杂因素，可以采用分层随机的方式进行分组，以达到控制混杂偏倚的目的。如前例，可先按是否吸烟进行分层，然后再在吸烟人群与非吸烟人群中分析饮酒与肺癌的关系，从而控制吸烟这个因素对研究结果的影响。

(2)配比设计。采用配比对照的方法将可疑混杂因素作为配比因素，使研究组与对照组在这些因素上保持一致。

(3)对研究对象进行限制。如考虑到年龄可能成为混杂因素，设计时可将研究对象的年龄限制在某一年龄段内；若性别可能是混杂因素，可仅以男性或女性为研究对象，以消除性别对研究结果的影响。

2. 资料分析阶段

(1)对资料进行分层分析。在资料收集之后，按可疑混杂因素进行分层，通过比较分层前后的效应值(*cRR* 与 *aRR*、*cOR* 与 *aOR*)来识别其混杂作用及其大小。

(2)利用标准化法(率的标化)进行调整。如欲比较甲、乙两县某因素暴露人群的发病率(生态性研究)，为消除年龄构成不同带来的影响，可对两县人口发病率按年龄进行标化，以消除年龄因素对研究结果的影响。

(3)利用多因素分析方法。采用多因素分析方法，将混杂因素与研究因素一并纳入分析。即通过一个数学模式把所有相关因素(包括多种混杂因子)对某一结局的影响进行定量综合分析。多因素统计分析方法包括多元线性回归，逐步回归，logistic 回归，Cox 风险比例模型等。

(李 革)

笔记

第四章 临床研究的常用设计方案

临床研究结果要转化为高质量的循证医学实践证据，最有效的方法之一就是研究之前进行严格的科研设计，尽可能地控制和减少非研究因素造成的偏倚，如患者的不同文化水平、经济地位，不同病程、病情程度，不同医疗护理水平等已知或未知影响因素。因此，选择与研究课题相适应的、论证强度高且可行性好的设计方案是保证临床研究质量的关键所在。

通常将临床研究的设计方案分为两大类：观察性研究（observational studies）和实验性研究（experimental studies）。观察性研究与实验性研究最主要的区别在于研究者不能人为地控制试验条件，不同比较组别系自然形成，只能尽可能控制非研究因素的影响，以求结论的真实、可靠。观察性研究进而又可分为分析性研究（analytical studies）和描述性研究（descriptive studies）等。实验性研究则可人为控制试验条件，如随机分组、根据研究目的设置合理的对照、盲法观测结果等，以探讨干预或治疗措施的真实效果。实验性研究也可分为临床试验（clinical trials）和类实验研究（quasi-experimental studies）等。相对而言，实验性研究的论证强度更高，结论更为可靠。表 4-1 列出了临床研究设计的常见方案及其证据级别。本章将分别介绍这些设计方案的概念、特点、设计模式、实施方案、资料的整理与分析、应用范围和优缺点，以便于大家在临床研究实践中选用。

表 4-1 临床研究设计的常见方案及其证据等级

分类	设计方案	证据等级
Ⅰ. 实验性研究		
Ⅰ(a)- 临床试验	随机对照试验	1a~2b
	半随机对照试验	2b~2c
	非随机同期对照试验	3a~4
	交叉试验	2c~3a
Ⅰ(b)- 类实验研究	自身前后对照试验	3a~3b
Ⅱ. 观察性研究		
Ⅱ(a)- 分析性研究	队列研究	1b~2b
	病例对照研究	3a~3b
Ⅱ(b)- 描述性研究	横断面研究	3b~4
	病例报告	4
	专家评述	5

* GRADE（Grading of Recommendations Assessment, Development and Evaluation）证据分级标准

第一节 随机对照试验

一、概 述

（一）定义

随机对照试验（randomized controlled trial, RCT）是采用随机化的方法，将合格的研究对象分别分配到试验组和对照组，然后接受相应的试验措施，在一致的条件下或环境中，同步地进行研究和观测试验的效应，并用客观的效应指标对试验结果进行科学的测量和评价。

笔记

(二) 历史回顾

17 世纪 F. Bacon 就提出应该“通过试验来尝试或证实特定疾病疗法的效果”。18 世纪 G. Berkeley 提出对照临床试验的观点。早期的流行病学实验主要是针对病人进行的，如 18 世纪 J. Lind 关于坏血病的治疗试验，19 世纪高木兼宽关于脚气病的治疗试验，以及 20 世纪初 J. Goldberg 关于燥皮病的治疗试验。“实验流行病学”最早就是指实验动物（大白鼠、小鼠）群的传染病流行模型研究，如 20 世纪 20~30 年代英国的 Topley、Wilson 和 Greenwood，德国的 Neufeld 以及美国的 Webster 等都曾报道了动物群的传染病（伤寒，肺炎）流行模型的经典研究，发现疾病流行同易感动物所占比例以及动物间接触程度有关。严格意义上的对照组（控制组）必须涉及随机分配，一般公认为 B. Hill 是随机对照试验之父。他设计的链霉素治疗肺结核效果评价方案，于 1946 年由英国医学研究会实施。1958 年美国开始有随机对照试验方法的临床研究报道。

近 50 多年来，随着理论和方法的日趋成熟，随机对照试验被公认为评价干预措施疗效的金标准或标准方案而广泛应用于临床研究中，为疾病治疗、预防和康复提供了大量真实、可靠的依据。

(三) 随机对照试验在临床研究和临床实践中的作用和价值

随机对照试验研究的新成果、新发现，极大影响着人们的临床实践活动。既往认为有效的治疗措施经随机对照试验证实无效。过去，某种治疗措施或药物是否应用于临床实践，主要取决于专家或顾问的意见。而临床医师选择治疗措施也主要基于个人既往治疗患者的成功经验或对疾病的病理生理学机制的理解。随机对照试验能够对某种治疗措施的真正疗效或副作用、影响因素进行科学的评判。

鉴于疾病发病机制的复杂性和对疾病认识的局限性，单纯根据疾病的病理生理机制、实验室研究结果推断某种干预措施在人体的疗效，有时可能误导。过去经病理生理机制研究认为有效的治疗措施，可能经 RCT 证实无效或有害。例如，从疾病的病理生理进程来看，心肌梗死患者发生室性心律失常是猝死的重要危险因素，因此有充足理由对此类患者常规使用抗心律失常药物，但随后的 RCT 证明，I 类抗心律失常药虽能抑制室性心律失常，但增加了病人的死亡风险；又如，短效钙拮抗剂尽管能有效降低高血压患者的血压水平，但心肌梗死和死亡风险反而上升。因此，要明确某种治疗措施短期和长期的临床疗效或副作用，必须进行以人体为研究对象的临床对照试验。

(四) 随机对照试验的准入条件

1. 遵循对照原则 由于生物学或医学的研究对象过于复杂，很难仅仅通过实验室或其他外在条件来控制干扰，因此需要将一部分研究对象设置为对照组或控制组（control group）。干扰作用（背景或基线）或自发变化体现在对照组上，通过与实验组比较，这些非处理因素的作用就可加以排除。

如设实验（处理因素）的效应为 T，干扰（非处理因素）的效应为 C，则：

实验（处理）组表现出来的混合效应 $=T+C$；对照组表现出来的效应 $=C$。则，实验（处理）的净效应 $=(T+C)-C=T$。

如果有两种处理（如实验处理和标准处理）的效应分别为 T1、T2，则：

处理 1 组表现出来的混合效应 $=T1+C$；处理 2 组表现出来的混合效应 $=T2+C$。则两处理组净效应差值 $=(T1+C)-(T2+C)=T1-T2$

这样，利用实验（处理）组与对照组的比较，从实验（处理）组的混合效应中分离出归因于处理的净效应。

需要注意的是：实验（处理）组表现出来的混合效应（T+C），是通过处理前后两轮测量值之差确定的。如处理为降血压药，给药前后两轮血压测量值之差，就是降血压药组表现出来的混合效应（T+C）。这里“降血压药组给药前”并没有起到所谓对照的作用，即体现出非处理因素效应 C。如果不存在非处理因素的干扰（C=0），那么降血压药的效应（T）就等于两轮血压测量值“自身

笔记

前后”比较之差，因而就不需要什么“对照”（控制）了，即成了“单纯实验设计”（simple experiment design）。这常见于严格控制的理化实验。因此，所谓“自身前后对照”的说法是不确切的。

2. 遵循随机化原则　随机对照试验中，采用随机化（randomization）的方法制订分配方案，并对分配方案进行隐藏，使合格的研究对象均有同等机会进入试验组或对照组，不以研究人员或研究对象的主观意愿为转移，可避免选择偏倚的干扰。在样本量足够大的情况下，可使若干已知 / 未知的或可测 / 不可测的混杂因素，在组间分布中维持相对均衡，从而有利于基线的可比性。如果样本量不够大，则不能保证混杂因素在组间均匀分布，导致基线不可比，此时若采用分层随机分组的方法，也可实现组间的均衡可比。

3. 遵循盲法原则　盲法（blinding，masking）指临床试验中，不让受试者、研究者或其他有关工作人员知道受试者接受的是何种处理（试验药物或对照药物），从而避免他们对试验结果的干扰。受试者若知晓自己接受的是何种药物后，会产生各种心理作用，影响对治疗效果的准确测量。研究或观察治疗结果者知道受试者接受的是何种处理，会自觉或不自觉地干扰他们对治疗效果的判定。资料的统计分析人员知道受试者接受的是何种处理，也可能影响他们对资料的取舍、整理或分析方法的选择等。根据盲法的使用情况，可以将临床试验分为开放试验、单盲试验、双盲试验和三盲试验。

（1）单盲试验（single blind trial）：指仅有受试者不知道自己是接受何种处理的试验。单盲法可以避免来自受试者方面的干扰。研究者（疗效观察者）知道处理分配情况，能及时做出用药调整，对患者安全有利，但不能避免来自研究工作人员方面的干扰。

（2）双盲试验（double blind trial）：指受试者和研究（疗效观察）者均不知道受试者是接受的何种处理的试验。当疗效和安全性指标受主观因素影响较大（如精神科量表）以及需要进行综合评价（总印象）时，必须使用双盲试验。国家食品药品监督管理局规定了应当使用双盲法的临床试验范围，这里的双盲法指受试者和研究工作人员“双方”处于盲态。

（3）三盲试验（triple blind trial）：指受试者、研究（观察）者和资料统计分析者均不知道受试者是接受的何种处理的试验。原理和具体实施同上述。

（4）开放试验（open trial）：指不用盲法的试验，受试者和研究者及有关工作人员都知晓受试者接受的是何种处理，如手术治疗与保守治疗的效果比较试验。

二、设计模式

（一）设计模式

随机对照试验的设计模式如图 4-1 所示。试验的研究对象必须采用公认的诊断标准加以确定，既可从患病群体（目标人群）中随机抽样，也可从住院或门诊患者中进行连续性非随机抽

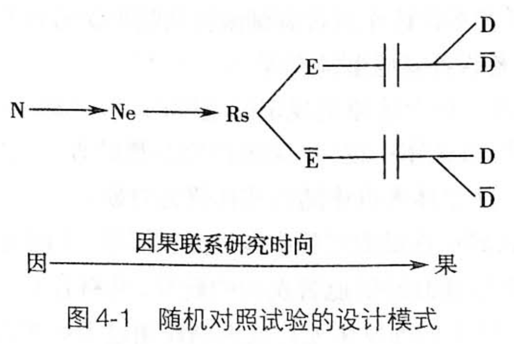


图 4-1　随机对照试验的设计模式

N：符合公认的诊断标准的病人总数或人群总数
Ne：该人群或病人中符合纳入标准又不具备排除标准的人数
Rs：随机分配或分层后随机分配
$\overline{E}$：未暴露可疑致病因素或未接受防治措施的对照组
E：暴露可疑致病因素或接受防治措施的试验组
┤├：为随访期或观察期，中间填时间
D：发病人数，有效人数，生存数等
$\overline{D}$：未发病人数，无效人数，死亡数等

笔记

样，再根据事先确定的纳入和排除标准，选择符合标准且自愿参加试验的患者，进而采用随机化方法将合格的研究对象随机分配入试验组或对照组，接受相应的干预措施，经过一段恰当的观察期后，测量治疗后的效果。根据结果的资料类型，采用相应的统计学方法进行分析、处理以评价干预措施的真实疗效及其组间差异。

（二）结果分析模式

根据图 4-1 的设计模式，可列出四格表（表 4-2），将试验组和对照组的结果分别填入相应的表格内，对两种干预措施的疗效进行分析和比较评价。

表 4-2 随机对照试验结果分析四格表

	结果		合计
	有效	无效	
试验组	a	b	$a+b$
对照组	c	d	$c+d$
合计	$a+c$	$b+d$	N

两组疗效比较可采用χ^2检验或 Fisher 确切概率法。

三、实施方案

（一）研究对象的选择

基于不同的研究目的，选择与之相适应的研究人群，在收集目标人群的基线资料的基础上，按照事先制定的、严格的纳入和排除标准来选择，以避免某些外来因素的影响。选择的主要原则有以下几点：

1. 应有纳入标准和排除标准 在明确的疾病诊断标准基础上，根据研究目的和具体条件，对受试者制定统一的纳入和排除标准。若纳入标准宽松，受试者代表性强，容易推论到临床实践，但受试者同质性差，所需样本量大，否则不易出现有统计学意义的结果。若纳入标准严格，则受试者同质性好，容易得到有统计学意义的结果，但推论应用范围受限。因此，在制定标准时需要综合考虑。

2. 选择干预措施有效且可以获益的人群 要考虑受试者的具体病情，使得他们既能从该临床试验中获得健康效益，又承担较小的风险。如研究抗心律不齐药物疗效时，最好选择近期频繁发作的患者，偶发患者不作考虑；又如评价某疫苗的预防效果，应选择某病高发区的易感人群为研究对象；再如评价强化补碘对轻度缺碘区居民的预防效果，宜选择孕妇和儿童。

3. 选择干预对其无害的人群 若干预对其有潜在伤害，不宜作为研究对象。这是医德伦理的基本要求。如有胃出血史者不宜选作对胃有刺激性药物的试验研究对象；而可能出现较大副作用的对象也不宜入选，如老弱者或危重病人等。

4. 选择有代表性的人群 在上述原则规定的范围内，应注意所选人群的代表性，如性别、年龄、种族、职业等特征是否与总体一致，以保证研究结果的外部真实性。为了保证研究对象的代表性，在可能条件下，应从总体人群中随机选择研究对象。

5. 选择能完成试验的人群 在试验过程中有可能被剔除、不能完成试验者不宜作为研究对象。例如用阿司匹林预防老年缺血性脑血管疾病的研究，常将合并有癌症者或有胃病者排除，因为这些人可能在研究尚未结束前即发生死亡或因副作用过大而被迫停止试验。另外，研究对象应有良好的依从性。

6. 应获得受试者的知情同意书 这是医学伦理的基本要求。

（二）样本量估计

为了控制Ⅰ、Ⅱ型错误率，在试验设计阶段，就应对研究所需的样本量加以估计。一般情况

笔记

下，样本量越大，Ⅰ、Ⅱ型错误率就越低；但样本量过大，不仅导致人力、物力、财力和时间的浪费，而且会给试验的质量控制带来诸多困难。

1. 影响样本大小的主要因素

（1）干预措施的效应大小：干预措施的效应越大，试验组和对照组的研究事件（疾病）的发生率的差异越大，所需样本量越小，反之，所需样本量越大。

（2）Ⅰ型（α）错误出现的概率：即出现假阳性错误的概率。确定的 α 越小，即要求的显著性水平越高，所需样本量就越大。通常将 α 定为 0.05。

（3）把握度（$1-\beta$）：β 为出现假阴性错误的概率（Ⅱ型错误率），$1-\beta$ 为把握度。要求的把握度越高，则所需样本量越大。通常将 β 定为 0.10。

（4）单侧检验或双侧检验：单侧检验比双侧检验所需样本量小。如果肯定试验组的效果好于对照组或只检验当试验组效果优于对照组时，就用单侧检验；否则就用双侧检验。

（5）研究对象分组数量：通常按公式估计的试验样本是指试验组和对照组分别所需的例数。如果依据研究目的要求设立了多个试验组或多个对照组，则每组都需达到估计的样本数。因此，分组数量越多，则所需总样本量越大。

2. 样本量计算

（1）计数资料的样本大小的估计：如果结局变量是计数指标，如发病率、感染率、病死率和治愈率等，则可按下列公式计算样本量：

$$N=\frac{\left[Z_{\alpha}\sqrt{2\overline{P}(1-\overline{P})}+Z_{\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}\right]^2}{(P_1-P_2)^2}\qquad \text{式(4-1)}$$

P_1：对照组某结局的发生率；P_2：试验组某结局的发生率；$\overline{P}$：$(P_1+P_2)/2$。

Z_{α}：为 α 水平相应的 u 值；Z_{β}：为 β 水平相应的 u 值。

N：为计算所得一个组的样本大小。

（2）计量资料的样本量估计：如果结局变量是计量资料，如血压、血糖等，则可按下列公式计算样本量：

$$N=\frac{2(Z_{\alpha}+Z_{\beta})^2\sigma^2}{d^2}\qquad \text{式(4-2)}$$

σ：为估计的标准差；d：为两组结局变量均值之差；Z_{α}、Z_{β} 和 N 所示意义同上述计数资料的计算公式。

（三）资料收集

在随机对照试验中，应对所有的研究对象进行随访。观察时间较短的试验，在随访终止时一次性收集资料即可；如果观察时间较长，则需要在整个观察期内分几次随访，其间隔以及随访次数视具体研究的需要而定。

1. 随访观察的内容　①干预措施的实施和标准化；②某些影响因素的变化，如饮食、其他疾病等；③有关结局或判断结局变量的各种临床和实验资料。

2. 随访观察的原则　①对试验组和对照组要采用同等的随访和资料收集方法；②对所有研究对象都要求随访到观察终止期，尽量避免失访或中途退出；③对随访调查人员应事先进行统一培训、统一资料收集的方法和标准。

3. 资料收集方法　①随访研究对象或知情人；②体检或采样进行实验室检测，例如测量血压、尿糖、血脂等；③到有关单位收集现成的资料，包括档案和记录等，如气象和环境监测资料、医院的病案、出生、死亡登记、工厂企业就业和工种档案、工作日志等；④环境调查，如居住及环境卫生情况、饮用水源、水质如何、工作环境如何等。

笔记

四、资料的整理与分析

(一) 均衡性检验

临床试验各比较组的可比性是评价研究结果的关键，因此资料分析的第一步要进行各比较组的均衡性检验，评价各种可能影响药物疗效、干预措施效果的因素在各比较组之间是否均衡，排除各种干扰因素对结果的影响，从而正确评价临床试验的结果。一般是采用 t 检验和方差分析检验计量指标在各组之间差异的显著性；采用 χ^2 检验、u 检验分析计数指标在各组之间差异的显著性。只有排除了干扰因素的影响，才能判断试验结果确为干预措施所起的作用。

(二) 统计描述和统计推断

就临床试验而言，其最终目的总是比较两组或多组疗效之间的差异。如果存在差异，则应检验其差异有无统计学意义，同时应评价其临床意义。临床意义的评价应包括对费用效益、费用效果、治疗副作用及治疗的复杂和方便程度等的评估。无论是治疗性还是病因学的 RCT 研究，结果分析内容主要为两(多)组计数指标的比较、两(多)组计量指标的比较、相关性分析以及多因素分析等方面。

1. 计数资料 如果结局变量是痊愈、好转、无效、死亡、生存等计数资料，则评价指标一般用率、比等相对指标，而不用绝对数，如治愈多少人等。在两(多)组治疗结果计数指标的比较分析中，主要的指标包括治愈率、有效率、不良反应发生率、病死率、病残率等。组间比较，可采用 χ^2 检验或 Fisher 确切概率法；若选用等级指标，如痊愈、有效、好转、无变化，可以进行 Ridit 分析；如果考虑多种因素对结果的影响，可以采用多因素分析的方法，如 logistic 回归分析等，这样既可得到研究因素(药物)的净效应，也可弄清有关因素的影响大小和方向。此外，还可以计算有关联系强度指标及其 95% 可信区间，如 *RR*、*AR*、*AR*% 等。近年来又陆续出现一些反映临床价值的系列指标，如 *RRR*、*ARR*、*NNT*、*NNH* 等，也可计算它们各自的结果及 95% 可信区间(参见有关统计学书籍)。随机对照试验中最常用的率有以下四种：

(1) 有效率

有效率 =(治疗有效例数 / 治疗的总例数)× 100%　　式(4-3)

注意治疗有效例数包括治愈人数和好转人数，治愈和好转需有明确的判断标准。

(2) 治愈率

治愈率 =(治愈人数 / 治疗总人数)× 100%　　式(4-4)

(3) 病死率

病死率 =(因某病死亡人数 / 某病患者数)× 100%　　式(4-5)

(4) n 年生存率

n 年生存率 =(活到 n 年的病人数 / 观察满 n 年的病人数)× 100%　　式(4-6)

生存率的计算还有 Kaplan-Meier 法、寿命表法等，请参见相关统计学书籍。

2. 计量资料 在结果为定量指标的比较分析中，主要指标为算术均数、几何均数及中位数等。两组间比较，可采用 t 检验；若是多组间比较，则可用方差分析，如果总体检验有统计学意义，还需要进一步做两两比较。若有多个观察测量时点，则应采用重复测量方差分析。同时对于结局变量是血压、血脂、血糖、血红蛋白等计量指标，除了仍可按照某些标准(如治愈、好转、无效)将其转换成计数资料处理外，可对两组每个对象治疗前后观察指标值的差(如血压下降多少)的均数进行比较。另一种方法是计算下降或升高的比例，如收缩压从 200mmHg 降至 140mmHg，可表示为下降 60mmHg，也可表示为下降了 30%。定量资料的分析要考虑资料是否满足正态性和方差齐性，若不满足，改用秩和检验；当考虑某一治疗性措施与结果的关系时，可做相关性分析，如判定不同药量、不同疗程、不同年龄等因素与疗效间的关系。同样在结果为定量指标的比较分析中可采用多因素分析方法，如多元线性回归，获得研究因素及其他因素对

笔记

结果的贡献大小。

（三）失访的处理

随机对照试验为前瞻性研究，无论随访时间长短，失访不可避免，失访率不同，对研究结果的影响也各有不同。建议采用意向治疗分析（intention to treat analysis，ITT），若只对完全依从的部分研究对象进行分析（per protocol，PP），结果可能高估。

五、应用范围

（一）临床治疗或预防性研究

随机对照试验最常用于治疗性或预防性研究，借以探讨药物、治疗方案、筛查方法等干预或预防措施的确切疗效，同时评价不良反应，排除非治疗和干预因素造成的毒副反应，为正确的医疗决策提供科学依据。

随机对照试验虽被公认为是最佳的治疗性研究设计方案，但并不能用于研究和解决所有的临床问题。在某些情况下，使用随机对照试验是不可行或不恰当的，如诊断性研究、疾病预后研究等。

（二）特定的病因学研究

多数情况下，随机对照试验并不适用于病因学研究，将某种致病因素和危险因素直接施加于人体，进行致病效应的研究是不符合医德的。例如，要了解吸烟在肺癌发病中的作用，人为设计一个随机对照试验，将原本不吸烟的研究对象随机分配入吸烟组或不吸烟组，随访数年，比较两组肺癌的发生率，显然既缺乏可操作性，也有违医德。但在特定的条件下，随机对照试验也可以用于病因学因果效应研究。应用的前提是：尚无充分证据证明某种可能致病因素对人体有危害，但又不能排除它与疾病的发生有关。若已有研究证明某一因素对人体有害，就不允许将该因素用于人体进行随机对照试验。

例如，妇产科为预防早产儿因缺氧带来的大脑损害和对今后智力发育不全的影响，曾对早产婴儿常规应用高浓度的氧气疗法。后来发现经此治疗的婴儿，出现了眼晶状体后纤维组织增生，导致不同程度的视力障碍，严重者甚至失明。经分析推论，认为可能与高浓度氧疗有关，为证实这种因果效应，于是采用了随机对照试验，一组早产儿继续用高浓度氧疗，另一组则用低浓度氧疗。经追踪观察分析，上述视力障碍确与高浓度氧疗有关，于是，临床上就淘汰了这一疗法。后来，这一病征被命名为“Terry's syndrome”（早产儿视网膜病变）。

（三）非临床试验的系统工程

随机对照试验还可应用于非临床试验的系统工程，如教育学和农业等领域。例如，要评价循证医学教育模式与传统医学教育模式的教学效果，可将条件相似的学生随机分配进入任何一组，课程结束后进行短期或长期效果的评估。

六、优缺点

（一）优点

1. 前瞻性的对照设计　可以人为控制研究对象的条件和暴露情况，对结果进行标准化评价；由于试验组和对照组是同步前瞻性的观察，故外部因素对结果影响较小，可以验证因果关系。

2. 可比性好　通过随机分组，将研究对象随机分配，特别是在某些情况下，将研究对象按影响结果的某些重要因素进行先分层、后随机，进入试验组和对照组，使各比较组的组间基线状况保持相对一致，可增加可比性。

3. 控制偏倚　采用随机分配和分配方案的完善隐藏，在选择和分配研究对象时可以较好地防止人为因素的影响，即使存在不为人知的干扰因素，也可维持在各比较组间的相对平衡，有效地控制了选择偏倚和信息偏倚。采取盲法评价疗效，避免了研究对象和疗效观察者的主观

笔记

因素所致的非特异性疗效和测量误差，可有效控制信息偏倚。

4. 诊断和实施标准化 对研究对象，采用严格一致的诊断/纳入和排除标准，有利于读者验证研究结果和确定研究结果的推广应用价值。同时实施过程采用明确的标准化的方法，能有效地控制各种干扰因素，使结果更趋真实。

5. 资料统计分析效能高 由于大多数统计方法是建立在随机抽样的基础上，因此随机对照试验更为适合，采用单因素和多因素的统计分析方法，能获得高质量的结果。

（二）缺点

1. 成本高 由于随机对照试验需要严格的设计，实施条件限制多，因此比较耗费时间，人力与财力支付较大。

2. 外部真实性受限 随机对照试验常因纳入/排除标准过于严格，虽使入选对象具有良好的同质性，但同时也导致其研究结果的代表性和外部真实性受限。

3. 医学伦理问题 在随机对照试验当中，如果不恰当地应用安慰剂、选用的对照组措施不当，或者让研究对象暴露于某种有害致病危险因素，或者研究对象不知情同意，都会出现医学伦理问题。

（黄悦勤）

第二节 随机对照试验的一些特例

一、交叉试验

（一）概述

交叉试验（cross-over design）是指试验中的试验组和对照组在整个试验过程中通过前后两个阶段互相交叉的方式完成，即分别先后接受两种不同试验措施的处理，最后评价试验结果的一种临床试验性研究。主要用于临床干预措施的研究和评价，是随机对照试验的一种特殊类型。它兼有随机对照试验和自身前后对照试验的优点，属于一级设计方案。

在试验开始前，符合纳入标准的试验对象先进入观察期，了解研究对象是否处于一个相对稳定的状态，例如在比较和评价新的治疗高血压的方案中，观察期的研究对象的血压已达正常水平，则不宜列入研究对象，以免影响研究结果的判断。经过一定时间观察后，将符合纳入标准的研究对象通过随机的方法分为试验组和对照组，分别接受第一阶段试验，如试验组首先接受方案A，而对照组接受方案B，经过一定时间试验观测，并获得相应的结果之后，按设计要求经过洗脱期（washout period）然后进入第二阶段的交叉试验。即两组进行交换，试验组换为方案B，而对照组换为方案A。按照第一阶段相同的指标，观测第二阶段的治疗反应。当试验结束时，将其结果与第一阶段的结果进行综合分析和评价。

在交叉试验设计时，一定要注意前后两个试验阶段药物消洗期的时间设置，通常以药物的5个半衰期为宜，不宜过短或过长，否则易受偏倚干扰的影响。

（二）应用范围

一般而言，交叉试验仅适用于慢性病且不易根治并需要药物维持治疗的某些疾病的研究。如高血压病、冠心病心绞痛、支气管哮喘等。因为这些疾病通常不用药物治疗就会发作。因此对于某些一治即愈者是不能采用交叉试验设计方案的。此外，在新药的开发和研究中，为减少样本含量的需要，Ⅰ期临床试验常采用交叉试验来观察药物的毒副反应，以便减少或消除个体间偏倚的影响。

（三）设计模式及特点

笔记

交叉试验通过随机的方法将受试者分成两组，甲组先接受方案A，乙组接受方案B，两组同

时进行观察。经过一定消洗期后，甲乙两组再交换接受方案 B 和方案 A。将两个不同阶段观察到的结果进行综合比较分析(图 4-2)。

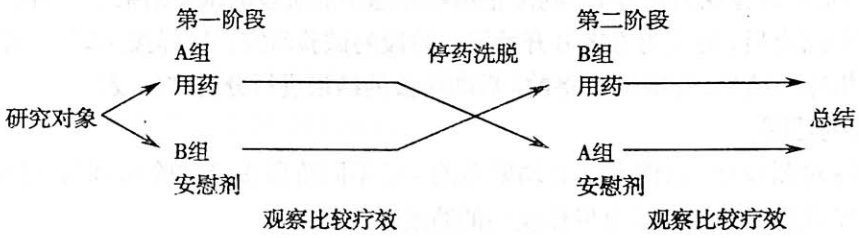


图 4-2　交叉试验设计模式图

(四) 结果分析

交叉试验中每个受试对象都要接受两种措施的处理，必然得到两种方案的观察结果。由于系自身前后对照，对定性资料应采用配对 χ^2 检验(McNemar 检验)，而定量资料则采用配对 t 检验和符号秩和检验。根据前后两方案的处理，将观察的结果分别列入配对四格表(表 4-3)，即方案 A 和 B 均有效的为 a，均无效的为 d，方案 A 有效而 B 无效的为 b，而方案 A 无效而 B 有效则为 c。

表 4-3　交叉试验结果分析四格表

方案 A	方案 B		合计
	有效	无效	
有效	a	b	$a+b$
无效	c	d	$c+d$
合计	$a+c$	$b+d$	N

(五) 优缺点

1. 优点

(1) 具有随机对照试验的优点：交叉试验作为随机对照试验的特例，采用随机分组、盲法测量和同期对照的方法，具有随机对照试验的全部优点，有效控制了选择性偏倚、信息偏倚和混杂偏倚，试验结果真实可靠。

(2) 设计特点：每个受试对象先后接受两种方案处理，具有自身前后比较的特点，消除了个体差异，同时也获得组间比较的结果，可大大减少样本量。

2. 缺点

(1) 应用范围受到限制，只适用于慢性、复发性疾病。

(2) 试验过程包括一定洗脱期，如过短则难以避免两种措施的沾染和干扰，过长影响试验周期，甚至使患者长时间得不到应有的治疗，影响预后。

(3) 如果试验周期长，容易发生失访、退出、依从性下降等问题。

(4) 每个病例在接受第二阶段治疗时，很难保证病情处于试验第一阶段开始时的相似状态，降低第二阶段的可比性，也影响疗效评估。

二、自身前后对照试验

(一) 概述

自身前后对照试验(before-after study in the same patient)指每一个受试对象，先后接受试验和对照两种不同措施进行试验研究，最后将两次先后观测的结果进行比较的一种设计方案。自身前后对照试验是以个体自身为对照，它可以避免个体差异对结果的影响。在研究过程中，试

笔记

验和对照两种措施的先后安排可以是随机的，也可以是非随机的，最好采用随机方法选择试验措施或对照措施作为第一阶段的试验。如方案A随机地进入第一阶段研究，受试者先接受方案A的干预，当完成试验观测任务后，则停止用药并总结前阶段的试验结果。然后进入消洗期，消洗期（洗脱期）结束后，更换为方案B开始第二阶段的试验研究。同样按照第一阶段方案A的测试指标观测相应的结果，完成后则将前-后两阶段的结果进行分析和比较。

（二）应用范围

自身前后对照设计，每例受试者均要在前-后不同阶段接受试验和对照两种措施，因此和交叉试验一样仅适用于慢性反复发作疾病的防治性研究。

（三）设计模式及特点

自身前后对照试验属于前瞻性研究设计，符合研究的纳入对象随机或非随机的在第一阶段接受一种措施的试验，然后经过一定的洗脱期后，受试者开始接受第二阶段的第二种措施。当完成试验后，将前-后的试验结果进行分析比较（图4-3）。

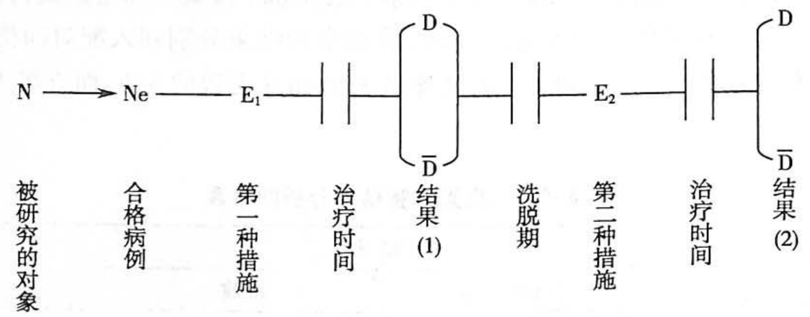


图4-3　自身前后对照研究设计模式

N：符合公认的诊断标准的病人总数或人群总数

Ne：该人群或病人中符合纳入标准又不具备排除标准的人数

$\bar{E}$：未暴露可疑致病因素或未接受防治措施的对照组

E：暴露可疑致病因素或接受防治措施的试验组

D：发病人数，有效人数，生存数等

⊣⊢：为随访期或观察期，中间填时间

$\bar{D}$：未发病人数，无效人数，死亡数等

（四）结果分析

参与自身前后对照试验的患者必须完成两个阶段的研究才能纳入结果分析，因此每例均有前后两种措施处理后的观察结果。受试对象可能有四种情况（表4-4），a为两种方案均有效，b为方案1无效而方案2有效，c为方案1有效而方案2无效，d为两种方案均无效。自身前后对照研究的结果属于配对资料，故定量资料采用配对t检验；而定性资料采用配对χ^2检验。

表4-4　自身前后对照研究结果分析四格表

第二种措施	第一种措施	
	有效	无效
有效	a	b
无效	c	d

（五）优缺点

1. 优点

（1）受试者以自身为对照，可消除个体差异，减少样本量，节约时间和成本。

（2）每例受试者均有接受试验和对照两种措施的机会，具有公平性。

（3）减少了自愿者偏倚和研究人员意愿偏倚。

（4）可以实现试验措施的标准化。

笔记

（5）试验中采用盲法并用随机方法安排前后干预措施，提高了结果可信度。

2. 缺点

（1）自身前后对照试验分前后两个阶段，难以保证两阶段的起始点完全一致，可能影响两个阶段的可比性。

（2）试验的应用范围有限，只适用于慢性、复发且不能自限自愈的疾病。

（3）洗脱期过长可能影响患者的及时治疗。

（黄悦勤）

第三节　其他类型的对照试验

一、半随机对照试验

半随机对照试验（quasi-randomized controlled trial）与随机对照试验设计相似，唯一的区别是试验对象分配方式不同。它是按半随机分配方式，即按研究对象的生日、住院日或住院号等末位数字的奇偶数，将试验对象分配到试验组或对照组，接受不同干预措施。

半随机对照试验常因分配方式，容易出现选择偏倚，造成基线情况的不平衡，因此，虽然花费的时间、精力、财力并不亚于随机对照试验，但其结果的真实可靠性却不及随机对照试验。

二、非等量随机对照试验

非等量随机对照试验（unequal randomized controlled trial）指试验对象按一定比例（通常为2∶1或3∶2）随机分配入试验组或对照组。主要应用于新药疗效验证研究，特别是由于病人来源和研究经费有限而研究者希望尽快获得结果的情况。但需要注意的是试验组和对照组的例数不能相差过大，否则会降低检验效能，超出1∶4或4∶1比例，检验效能会明显下降。

三、整群随机对照试验

整群随机对照试验（Cluster randomized controlled trial）不同于多数随机对照试验以患者个体为随机分配单位，而是在某些特殊情况下，以多个个体组成的小群体作为分配单位，进行随机分组。

例如，设计一种预防心血管病的特殊膳食食谱，拟与普通饮食比较，观察预防心血管疾病发生的效果。假若一家4口人中，有3个被选为合格的试验对象，其中有可能1人分配到特殊膳食组，2人分配到普通饮食组。在日常生活中，一家人不可能长期做两种不同膳食供用，即使做到了，沾染或干扰现象也不可避免，从而影响研究结果。显然单个体不宜作为试验的分配单位。于是，就可以一个家庭，一对夫妇，一个小组甚至一个乡镇等作为随机分配单位，将其随机地分配入试验组或对照组，分别接受相应的措施，进行研究。此类试验称为整群随机对照试验。

整群随机对照试验在设计上与一般随机对照试验一样，不同之处在于因随机分配的单位不同，导致样本含量和结果分析方法有所差异，所需样本含量较大。

四、基于单个患者的随机对照试验

单个患者的随机对照试验（number of one randomized controlled trial，n-of-1 trial）是基于罹患慢性疾病的单个个体，进行的一种随机对照试验，以确定多种治疗措施中哪一种对该患者有效，从而避免服用多种药物，减少浪费以及避免误服某些无效甚至有害药物。

方法是将所有“有效”的药物与其安慰剂配对，以每对药物为一个单位，采用随机分配的方

笔记

式决定每对药物的使用顺序；对每对药物，同样以随机分配的方式决定试验药物和安慰剂的使用顺序。依药物疗效发生和达到稳定所需的时间来决定药物的观察期，所有试验药物的观察期应保持一致，以便比较。通常采用双盲法，以利于试验结果的评价。基于单个患者的随机对照试验适用于慢性复发性疾病。

五、非随机同期对照试验

非随机同期对照试验（concurrent nonrandomized controlled trial）是临床传统使用的一种研究设计，指试验组和对照组的受试对象不是随机分配的，而是由病人或医生根据病情及有关因素人为地分到试验组或对照组，并同期进行结果观察。

非随机对照试验是前瞻性的研究。常用于比较临床不同干预措施的实际效果。该试验在研究对象的分组分配上，由于人为的因素，往往会造成试验和对照两组之间在试验前即处于不同的基线状态，缺乏可比性。在研究过程中也难以盲法评价试验结果，使得许多已知 / 未知的偏倚影响观测结果的真实性。但在临床实际工作中，有些情况下不适宜做随机对照试验，例如外科手术治疗、急重症病人抢救或贵重药物的选用等。因此，只能根据具体情况将患者分入试验组或对照组。其研究结果的论证强度虽远不及随机对照试验，但在尚无随机对照试验结果或不能进行随机对照试验时，还是可取的，尤其是对于大样本的非随机同期对照试验结果，仍具有重要的临床价值。只不过在分析和评价研究结果的价值及意义时，应持审慎的科学态度。该研究方案的可行性好，易为临床医生和患者接受，依从性较好；但由于选择性偏倚和测量性偏倚的影响，降低了结果的真实性，其论证强度也相应减弱。

六、多中心临床试验

（一）设计原则

多中心临床试验（multicenter clinical trial）是由多位研究者按同一试验方案在不同地点和单位同时进行的临床试验；各中心同期开始与结束试验；多中心试验由一位主要研究者总负责，并作为临床试验各中心间的协调者。大型多中心合作临床试验一般均有 1000 例以上大样本。

大型多中心合作的临床研究常有以下两种：一种是大样本随机临床试验，一种是Ⅲ期新药临床试验。两者均是为了评估某种治疗措施的临床效果，但也有不同之处。大样本随机临床试验是医疗科研人员发起的、旨在解决医学领域某些尚待解决的问题进行的临床研究，用以评估某种治疗措施对患者生存率及重要临床事件的影响；而Ⅲ期新药临床试验是药品生产厂家按照药品法规定，为达到新药注册的目的所进行的试验研究，旨在评估该药的临床疗效及不良反应。

进行大型多中心合作临床试验，从社会效益方面看，因其设计方法科学，可公正地评估治疗药物的疗效，限制或淘汰疗效差或有严重不良反应药物，能为循证医学实践提供证据，有助于提高临床治疗水平；从经济效益方面看，进行大型多中心合作临床试验，既可为国家和患者节省大量医药开支，又可让安全有效药的生产厂家获益，从而实现双赢。此外，大型多中心临床试验也是上市后药物疗效再评价的最佳方法。

（二）实施方法

1. 统一方案 多中心合作临床试验的设计和实施必须考虑到多中心的特点，研究方案要由各参与方的主要研究者共同讨论以后确定，然后以书面方式予以确认。

2. 医学伦理 研究方案及其附件要由伦理委员会讨论通过并做出书面同意后方能执行。多中心合作临床试验涉及一个以上研究机构的伦理委员会。

3. 统一培训 多中心合作试验同时涉及多个研究机构，而每一个研究机构，除主要研究者外，还有诸多研究参与者。为使他们能够按同一标准来执行研究方案中的每一个具体细节，需要对研究参与者进行规范化培训，该临床试验的所有参与人员，如医师、护士、药师、检验人员，

笔记

均应进行培训。

4. 研究的同步性 多中心合作临床试验要在各研究机构同步进行，因此应规定各个研究机构中首例受试者入组时间和最后一名受试者入组及完成时间。这可使临床试验统一在规定时间内完成，不同中心完成时间若相差过大，可能会导致中心效应。

5. 评价 多中心合作临床试验中应采用统一的安全性和疗效评价方法，包括实验室检查和临床检查方法，如从血、尿常规，到血生化、肝肾功能检测，乃至X线、心电图以及特殊的形态和功能检查等。同一个检查项目若不同实验室采用不同的材料和方法，其检测结果就难以比较。为此，建议在临床检验方面，采用中心实验室的办法，即为多中心合作临床试验而专门建立的实验室，其各项检查均采用国际公认方法，所用试剂质量可靠，操作规程透明、质量控制明确，同时权威机构应定期进行质量稽查。此外，还应建立一套标本的采集、转运、储藏体系，将各中心的标本集中到中心实验室进行检验，统一出检验报告。

6. 数据管理 各中心在试验时产生的大量数据，应传送到一个数据处理中心进行统一管理，包括查询、核对、储存，最后进行分析。

7. 监督 在多中心合作临床试验中，要配备专职的监察员。监察员的职责是负责试验研究的质量监控，旨在保证各中心严格按方案执行，并对严重违背方案者及时上报。因此，监察员首先应经过严格的《药品临床试验管理规范》(Good Clinical Practice，GCP)培训，熟悉相关法规，具有一定的临床试验经历或经验。同时要求掌握相关专业临床知识，具有识别原始资料和证据的能力，另外监察员还应熟悉本项目的研究方案及相关文件，工作认真负责，有足够精力进行定期的监察工作等。监察过程分为三个阶段，即监察前、监察中和监察后。监察内容具体包括协作医院的科研能力及研究者对试验方案的掌握，中心随机化系统的使用情况，研究者对方案的执行情况，药品的储存、供应与回收，各研究单位向项目办公室的资料传真情况，对试验数据真实性的考察，主要观察指标的记录情况，以及了解研究者对该项试验的意见等。

8. 管理规范 大型多中心合作临床试验的实施必须有一套合理的科学管理规范，涉及临床试验设计方案及三大原则强化、人员培训、组织机构构建，以及药品管理、医学伦理、资料管理、定期监测、终点事件评估、质量控制、监察员职责乃至资料统计分析和总结报告等诸多环节或内容，均需要加以规范。

9. 科研经费预算 预算管理同样需要统筹规划、细心组织。科学、规范地编制综合预算，是预算管理制度改革的一项重要内容。预算编制要紧紧围绕资金收支两条线，每一收支项目金额应有充分翔实的依据说明，项目计算应细致认真，不能随意编造。确保预算能够顺利贯彻执行，同时执行中应做到有效控制、信息及时反馈、实施严格监督和考核制度。

课题研究经费指的是一旦课题确立后，课题研究过程中所产生的支出费用，包括与课题研究有关的所有直接费用、间接费用和协作研究费用。

10. 资料管理和分析 大型多中心合作临床试验要求按照研究方案对数据进行统一管理，包括数据录入、数据核查、数据储存以及数据分析。对于多中心试验的数据管理，应有专门负责人，其职责是全程参与研究方案、病例记录表(CRF)的设计，提出统计分析要求，定期访问各试验分中心，监控数据收集质量。统计分析人员也应全程参与研究方案设计，负责制订统计分析计划和撰写统计分析报告，同时还要参与撰写试验报告和论文等。

(三)项目管理

1. 基本概念 项目是指一次性、多任务的工作，明确规定了项目开始和结束日期、特定的工作范围、预算和拟达到的特定性能水平等。项目管理是指：在项目活动中通过运用专业知识、技能、工具和方法，使项目能够实现或超过项目干系人的需要和期望。这一定义不仅强调使用专门知识和技能，还强调项目管理中各参与人的重要性。因此，项目管理就是为实现目标而在项目执行过程中进行计划、安排进度、控制、沟通和激励，旨在取得良好效益的各项活动统称。

笔记

项目管理包括启动、计划、执行、控制、收尾五大环节。

大型多中心合作临床试验作为一种医学研究项目，同样需要遵循项目管理的一般原则。在多中心合作临床试验中需要进行质量管理的包括中央实验室（centralized laboratory）、数据获取和报告（data acquisition and reporting）、（远程）数据输入[（remote）data entry]、病例记录表系统（case report form systems）、临床数据管理（clinical data management）、不良事件报告（adverse event reporting）、临床供给系统（clinical supply systems）、统计分析系统（statistical analysis systems）等，均涉及专门的知识和技能。

同时，多中心临床试验要求各参与方分工协作、共同完成。参与人员包括临床研究人员、统计专业人员、计算机程序分析员（编程人员）、数据输入录码人员等。试验过程中强调各参与人的重要性，符合项目管理的范畴。

2. 项目管理内容

（1）项目范围管理：是用以保证项目包含且只包含所有需要完成的工作，以便顺利完成项目所需要的各个过程，由范围计划、范围定义、范围核实和范围变更控制等构成。

（2）项目时间管理：是用以保证能够按时完成项目所需的各个过程，由活动定义、活动排序、活动历时估算、进度计划编制和进度计划控制等构成。

（3）项目成本管理：是用以保证在批准预算内完成项目所需要的各个过程，由资源计划、成本估算、成本预算和成本控制等构成。

（4）项目质量管理：是指用以保证项目满足其所执行标准的要求而所需要的各个过程，由质量计划、质量保证措施和质量控制等构成。质量计划是质量管理的起始点，质量计划需要反映项目质量管理中的各个方面。

（5）项目人力资源管理：是用以保证参加项目的人员能够被最有效使用而所需要的各个过程，由组织的计划编制、人员获得和队伍组建等构成。项目人员由管理人员和实施人员两部分组成，管理人员是人员组成中的关键，也是项目管理的主要实施人。

（6）项目沟通管理：是用以保证项目信息能够被及时、正确地产生、收集、发布、储存和最终处理而所需要的各个过程，由沟通计划、信息发送、绩效报告等构成。

（7）项目风险管理：是有关识别、分析和应对项目风险的各个过程，由风险管理计划、风险识别、风险分析、风险应对和风险监控等构成。

（8）项目采购管理：是用以从执行机构以外获得物资和服务的各个过程，由采购计划、询价计划、供方选择和合同管理等组成。

（9）项目文档管理：临床试验完成后的档案资料是必不可少的，试验过程中的各种会议纪要是质量控制和合同管理的重要组成部分，是试验完成后充分利用资料的根本保证。

（黄悦勤）

第四节 队列研究

队列研究（cohort study）是探讨疾病病因的常用方法之一，其论证强度较高，能较好地揭示两事件间客观存在着的因果关系。而其设计原理也常常用在疾病预后（见本教材第十一章）、临床试验等研究中。

一、概 述

（一）概念

队列研究是将一群（组）研究对象（队列）按是否暴露（exposure）于某研究因素分为暴露组与非暴露组（对照组），随访观察适当长的时间，比较两组之间所研究疾病（或事件）的发病率（或

笔记

发生率）或死亡率差异，从而判断这个（些）暴露因素与疾病之间有无关联及关联大小的一种观察性研究方法。队列研究又被称为群组研究、定群研究、前瞻性研究（prospective study）等，但以队列研究这一名称最常用。"暴露"是指研究对象接触过某种因素（如重金属），或具有某些特征（如年龄、性别、遗传）、行为（如吸烟）。暴露可以是危险因素，也可以是保护因素。

（二）特点

1. 属于观察性研究，队列研究中所观察的暴露因素不是人为给予的，而是客观存在于研究人群中，研究因素的暴露情况及其变化是由研究者观察获得的。

2. 属于"前瞻性"研究，研究开始时，研究的疾病尚未发生，或观察的临床结局尚未出现，随访一段时间后，才能观察到研究的结局是否发生。

3. 研究对象按是否暴露于所研究的因素分为暴露组和非暴露组，其暴露状况已客观存在，研究者不能将其随机化分配。

4. 由"因"推"果"，符合先因后果的推理逻辑。

二、设计模式

队列研究设计的基本原理见图 4-4。队列研究按其研究时间的起止点（时序），又可分为三种设计模式：前瞻性队列研究、回顾性队列研究和双向性队列研究。

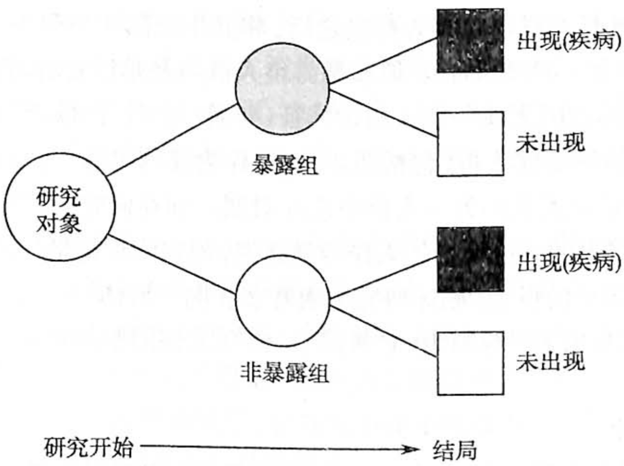


图 4-4　队列研究基本原理示意图

1. **前瞻性队列研究**（prospective cohort study）　是指观察时间从现在开始，追踪观察到将来某个时间，了解其发病或死亡情况，以确定某暴露因素与疾病的关系。通常提到的队列研究就是指这种研究，是队列研究的基本形式。

2. **回顾性队列研究**（retrospective cohort study）　是指以过去某个时间为起点，收集基线和暴露资料，以当时人群对研究因素的暴露情况将其分为暴露组和非暴露组，追踪观察到现在发病或死亡的结局情况，以研究暴露与疾病的关系。这种设计模式又称为历史性队列研究（historical prospective study）。回顾性队列研究的前提是过去有关暴露与发病的记录必须准确和完整。尽管收集暴露与结局资料的方法是回顾性的，但究其性质而言仍是从因到果的研究方法。

3. **双向性队列研究**（ambispective cohort study）　是指在回顾性队列研究之后，继续追踪观察到将来某个时间，又称为混合型队列研究，它是前瞻性队列研究和历史性队列研究方法的结合，兼有上述两种队列研究的优点，并在一定程度上弥补了相互的不足。在进行回顾性队列研究的过程中，如果从暴露到现在不能达到观察结果所需的足够观察时间，还需继续前瞻性观察一段时间时，就可选用双向性队列研究。

笔记

三、实施方案

在确定开展某因素(如:吸烟)和结局(如:肺癌)因果关联的队列研究后,实施方案的制订是研究成败的关键一环。

(一)研究对象的选择

1. 研究现场 由于大多队列研究的随访时间较长,因此研究现场必须有足够的人口数量,人口相对稳定,当地的领导重视,群众理解和支持,愿意合作,当地也要有较好的医疗条件,交通便利,便于随访。人口流动大的地区或单位,失访率会较高,对结果有较大影响,一般不适于作为研究现场。

2. 研究人群 研究人群包括暴露组和非暴露组,根据研究目的和研究条件的不同,研究人群有不同的选择方式。

(1)暴露组:即对研究因素有暴露的人群。通常可以选择在某社区或地理区域内居住的全体人群,其中暴露于某研究因素(如:吸烟)的人群即为暴露组。如果要研究一些特殊暴露、职业暴露与疾病的关系,往往要选择特殊或职业人群,因为这些特殊暴露人群某种疾病的发病率或死亡率高于其他一般人群,有利于探索该暴露因素与疾病之间的联系。如选择石棉作业工人研究石棉与肺癌的关系等。职业人群有关暴露与疾病的历史记录较为全面、真实和可靠,因此常作为历史性队列研究的暴露人群。

(2)非暴露组:即对照人群。观察人群确定后,将其中暴露于所研究因素的对象作为暴露组,其余即为非暴露组(如:不吸烟者)。但非暴露组人群与暴露组要有可比性。对照人群除未暴露于所研究因素外,其他因素或一般人口学特征(年龄、性别、民族等)都尽可能地与暴露组人群相同。但是,当以特殊暴露人群(包括职业人群)作为暴露组时,不应在同一人群选择对照组,而应在与该暴露或职业无关的另一人群中选择对照。如在研究二硫化碳(CS_2)与冠心病的关系时,就以有 CS_2 暴露史的纤维厂的工人作为暴露组,而以附近造纸厂的工人作为非暴露组。从中可看出,对于非暴露组的设置,临床研究中通常多在同一群体中进行,谓"内对照";若研究某种环境因素的致病效应时,则对照组(非暴露组)应在无该因素的地区或人群中选择,谓"外对照"。

(二)样本量的估计

1. 样本量估计有关的参数 估计样本含量之前,必须确定下述参数:

(1)非暴露人群或全人群中被研究疾病的发病率(P_0):可通过查阅文献或预调查获得。P_0 越接近0.5,所需样本量越大。

(2)暴露人群中的发病率(P_1):可通过查阅文献或预调查获得。如果已知预期相对危险度(RR),则可通过公式 $P_1=RR\times P_0$ 求得。P_1-P_0 差值越大,所需样本量越小。

(3)显著性水平(α 值):是检验假设时的Ⅰ型错误(假阳性错误)α 值。假阳性错误出现的概率越小,所需样本量越大,一般情况取 $\alpha=0.05$ 或0.01。

(4)把握度($1-\beta$):又称为检验效能(power),反映能够发现疾病与病因之间确实存在关系的概率。β 是检验假设时Ⅱ型错误(假阴性错误)的概率,把握度越高,β 值越小,所需的样本量越大,通常取 $\beta=0.10$ 或0.20。

2. 样本量估算方法 估计样本量,常用两种方法:查表法与公式法。查表法根据已知的上述 α、β、P_0 和 RR 四个基本参数,可从参考书的相应附表上查出。公式法如下所述:

在暴露组和对照组样本量相等的情况下,可用下式计算出各组所需的样本量。

$$n=\frac{(Z_\alpha\sqrt{2\overline{P}\,\overline{Q}}+Z_\beta\sqrt{P_0Q_0+P_1Q_1})^2}{(P_1-P_0)^2}$$

笔记

式中：P_1 和 P_0 分别代表暴露组和对照组预期发病率，$\overline{P}=(P_0+P_1)/2$，$\overline{Q}=1-\overline{P}$，$Q_0=1-P_0$，$Q_1=1-P_1$，$Z_\alpha$、$Z_\beta$ 为标准正态分布下的面积。

在计算样本量时一般按 10%~20% 估计失访率，所以应在原估计样本量的基础上加 10% 以上作为实际样本量。

（三）资料的收集与随访

1. 收集基线资料　队列人群确定后，应全面收集每个研究对象在研究开始时的基本信息，即基线资料（baseline information），这些资料是区分暴露组和非暴露组，判定研究结局的重要依据。资料收集的方法包括：①调查询问：调查的内容包括一般人口学特征（如性别、年龄、职业、民族、居住地、经济收入等）、研究因素的暴露情况（如吸烟、饮酒等）及身体健康状况等。②查阅现成记录：如常规的出生死亡登记，各种人口与疾病统计，医院病历，户口登记等。③体格检查和实验室检查：对于身高、体重、血压、血脂、血糖等研究相关指标需做相应的检查。④环境调查和检测等。

2. 随访观察　暴露因素确定后，结局事件的发生往往需要一定的时间，在这个过程中，就必须对研究对象进行随访（follow up）。并要根据研究目的和具体病种确定随访间隔时间和终止时间。

随访对象是暴露组和非暴露组的所有研究对象。在相同时期内以相同的方式对两组人群进行随访。绝对不可有先有后，中途放弃或漏访。随访方法与基线资料的收集方法相同，值得注意的是，随访方法在整个随访过程应保持不变。

随访内容依研究目的和设计要求不同而各异。一般而言，应与收集基线资料的内容和方法一致。但是应重点观察以下内容：①暴露人群暴露情况及程度有无改变。如吸烟者吸烟量有无变化，是否戒烟等；对某些定量指标如血压、血糖等进行体格检查和实验室检查。②收集结局相关资料。如发病日期、入院时间、诊断方法、死亡原因、死亡时间和地点等。③人口变动的情况。如进入、退出、失访人数等。

3. 观察终点　随访观察终点（end-point）多为发生疾病或死亡，但必须是研究所限定的疾病或死亡。观察过程中发生其他疾病或死亡不能视为结局。如研究高血压和冠心病的关系，则观察终点是冠心病，随访过程中，研究对象发生了肿瘤或其他疾病，则不能认为是到了观察终点，应继续随访；此时如果某研究对象死于肝癌，应作为失访处理，不能视为结局。终点指标与测量标准都应在研究设计阶段做出明确规定，在观察中途不能改变，对全部人群应按事先制定好的诊断标准统一进行评价，否则将造成结果的偏倚。

队列研究观察时间较长，在随访过程中应加强质量控制，以保证研究结果的可靠性和准确性。

四、资料的整理与分析

队列研究资料主要是计算暴露组和非暴露组的发病率或死亡率，检验其差异的显著性，分析暴露因素与发病或死亡之间是否有联系。如存在联系则进一步计算相对危险度等指标，分析联系的强度。

队列研究资料整理的基本形式见表 4-5。表中暴露组的发病率为 $a/(a+b)$，非暴露组的发病率为 $c/(c+d)$，在控制各种偏倚后，如果暴露组的发病率显著高于非暴露组且有统计学意义，则说明暴露因素与该疾病有一定关联，且很可能是因果关系。

表 4-5　队列研究资料整理表

组别	病例	非病例	合计
暴露组	a	b	$a+b$
非暴露组	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d$

笔记

1. 率的计算　队列研究计算的发病概率与普通发病率不同，根据队列人群的数量、稳定程度、发病强度和观察时间的不同，可以计算累积发病率、发病密度等指标。

（1）累积发病率（cumulative incidence，*CI*）：当研究对象人数较多，人口稳定，观察时间较短，资料比较完整，且失访人数较少时，以该研究人群观察开始时的人口作分母，以观察期内某病的新发病例数作为分子计算累积发病率。累积发病率的量值变化范围为 0~1，其数值高低受研究时间长短的影响，报告累积发病率时应说明观察时间的长短。

累积发病率的基本计算公式为：

$$n\text{年的某病累积发病率}=\frac{n\text{年内的新发病例数}}{n\text{年内的平均暴露人口数}}\times 1000‰（100\,000/10\text{万}）$$

（2）发病密度（incidence density，*ID*）：当研究人群观察时间较长，研究对象不断增减、队列人群难以稳定时，则需要以人时（person-time，*PT*）为单位代替平均人口数作分母来计算率。这时计算出来的是一种人口数有“变动”的发病率，即发病密度，而不是静止的累积发病率。计算发病密度时，分子为观察期间研究疾病的发病数，分母则不是普通的人口数，而是用人时来表示。人时等于观察人数乘以随访时间，人时单位可为人年、人月等，通常用人年来表示。如 1000 人观察了 1 年或 100 人观察了 10 年均为 1000 人年。这种以人时计算的发病密度又称为人时发病率。在比较几项随访时间不等的前瞻性研究的发病率时多用发病密度。

发病密度的计算公式如下：

$$\text{发病密度}=\frac{\text{某人群在观察期内的发病数}}{\text{观察期内的观察对象人年数}}\times 100\,000/10\text{万}$$

计算人年的常用方法有近似法和寿命表法。

2. 联系强度的估计　当统计学检验提示暴露与疾病具有显著的统计学意义时，应进一步计算联系的强度。常用的指标有以下几种。

（1）相对危险度（relative risk，*RR*）：又称率比（rate ratio），是暴露组发病（或死亡）率（I_e）与非暴露组发病（或死亡）率（I_o）的比值。它表明暴露组发病或死亡的危险性为非暴露组的倍数。其计算公式为：

$$RR=\frac{I_e}{I_o}=\frac{a/(a+b)}{c/(c+d)}$$

$$RR\text{的 95\% 可信区间}=RR^{1\pm1.96/\sqrt{\chi^2}}$$

在队列研究中 *RR* 值是反映暴露与疾病联系强度的重要指标之一，反映了暴露于某因素所致的发病（或死亡）率是不暴露者的倍数，具有病因学意义。$RR<1$ 说明暴露因素是疾病的保护因素，$RR>1$ 说明暴露因素是疾病的危险因素，$RR=1$ 说明暴露因素与疾病无联系。

（2）归因危险度（attributable risk，*AR*）：又称特异危险度或率差（rate difference）。是指暴露组的发病率或死亡率（I_e）与非暴露组的发病率或死亡率（I_o）之差。*AR* 表示暴露者中完全由某暴露因素所致的发病率或死亡率，公式如下：

$$AR=I_e-I_o$$

$$\text{或 } AR=I_o(RR-1)$$

$$AR\text{的 95\% 可信区间}=AR^{1\pm1.96/\sqrt{\chi^2}}$$

AR 说明暴露人群比非暴露人群增加的疾病发生率，如果去除该因素，就可减少这个数量的人发病，因此更具疾病预防和公共卫生意义。

（3）归因危险度百分比（attributable risk percent，*AR*%）：表示暴露者中由暴露所致的发病率或死亡率（I_e-I_o）占暴露者发病率或死亡率（I_e）的百分比。*AR*% 与 *RR* 的大小有关。公式如下：

笔记

$$AR\%=\frac{I_e-I_o}{I_e}\times 100\%$$

$$或\ AR\%=\frac{RR-1}{RR}\times 100\%$$

（4）人群归因危险度（population attributable risk，*PAR*）：表示整个人群中，暴露因素所引起的发病率增高的部分，又称为病因分值（etiologic fraction）。是整个人群中该疾病的发病（或死亡）率（I_t）与非暴露人群发病（或死亡）率（I_o）之差。表示人群中因暴露于某因素所致的发病（或死亡）率。其计算公式为：

$$PAR=I_t-I_o$$

（5）人群归因危险度百分比（population attributable risk percent，*PAR*%）：表示人群中由于暴露所致的发病率或死亡率占人群总发病率或死亡率的百分比。*PAR*% 在卫生行政部门制订疾病预防策略时需加以考虑，也可用于疾病预防的宣传教育，它可向群众说明在完全控制该暴露因素后人群中某病发病（或死亡）率可能下降的程度。其计算公式为：

$$PAR\%=\frac{I_t-I_o}{I_t}\times 100\%$$

五、应用范围

1. 疾病预后研究　队列研究可以观察人群暴露于某因素后，疾病逐渐发生、发展，直至结局的全过程，包括亚临床阶段的变化与表现，因此队列研究不仅可以了解个体疾病的自然史，而且可了解疾病在整个人群中发展和流行过程。队列研究，特别是前瞻性队列研究成为疾病预后研究的首选设计方案。

2. 检验病因假设　检验病因假设是队列研究的主要用途和目的。一般说来，队列研究往往是明确因果联系最有力的方法。

3. 评价预防效果　发现暴露因素对预防疾病常常具有明显的作用。如戒烟可减少既往吸烟者发生肺癌的危险，那么对戒烟效果的评价就是队列研究的目的。

4. 新药上市后监测及疗效比较研究（comparative effectiveness research，CER）　新药通过三期临床试验上市后，还需要对其不良反应等进行长期监测，队列研究可以更大样本和更长时间地观察该新药的各种不良反应。

此外，在疗效比较研究中，也可进行队列研究，如大型临床注册研究（patient registry study）等。

六、优缺点

1. 优点　符合病因链先因后果的时间顺序，验证病因与疾病之间的因果关系论证强度高；可以直接计算暴露组和对照组的发病率或死亡率，获得 *RR*、*AR* 等指标；一般不存在回忆偏倚；可以了解疾病的自然史；能对暴露因素所致的多种疾病同时进行观察，从而获得一种病因与多种疾病的可能因果关系。

2. 缺点　所需研究时间长，样本量大，人力、物力投入大，容易产生失访偏倚，通常不适用于发病率低、潜伏期长的疾病病因研究。

（闫永平）

第五节　病例对照研究

病例对照研究（case-control study）是临床上开展病因研究最具实用价值的一种设计方案，它通过科学合理的对照设置，可以在一定程度上防止偏倚的干扰，是探讨病因及危险因素的最为

笔记

常用研究设计之一，同时对于治疗效果和预后等方面的研究也可选用此方案，而且随着病例对照研究方法的不断完善，其应用范围也在不断扩大。

一、概　述

（一）概念

病例对照研究是选择一组患有所研究疾病的人作为病例组，选择一组不患有所研究疾病的人作为对照组，调查这两组人对某个（些）因素的既往暴露情况，通过比较两组间暴露率或暴露水平的差异，用以判断该疾病与这个（些）因素的关系。因为这种研究方法是比较病例组与对照组既往的暴露史，在时间上是"回顾性"的，故又称为回顾性研究（retrospective study）。

（二）特点

1. 按发病与否分成病例组与对照组，病例对照研究是在疾病（事件）发生后进行的，此时已有一批可供选择的病例。然后再选择一组无所研究疾病的人作为对照组。

2. 调查的暴露情况是由研究对象从现在对过去的回顾。也就是说，我们关注的是研究开始之前，病例组和对照组对所要研究因素的暴露情况。

3. 由"果"推"因"，研究中是先有结果，即已知研究对象患某病或不患某病，再追溯其可能与疾病有关的原因。

4. 病例对照研究受到回顾性观察方法的限制，不能观察到由"因"到"果"的发展过程并证实其因果关系。只能通过两组暴露率的比较来分析暴露与疾病是否有关联。

二、设计模式

（一）病例对照研究

病例对照研究的基本原理见图 4-5。若病例组某因素的暴露率或暴露水平明显高于对照组，且研究过程又无明显的偏倚，则该因素或措施与所研究的疾病有联系。病例对照研究可分为成组病例对照研究（group case-control study）和配对病例对照研究（matched case-control study）。

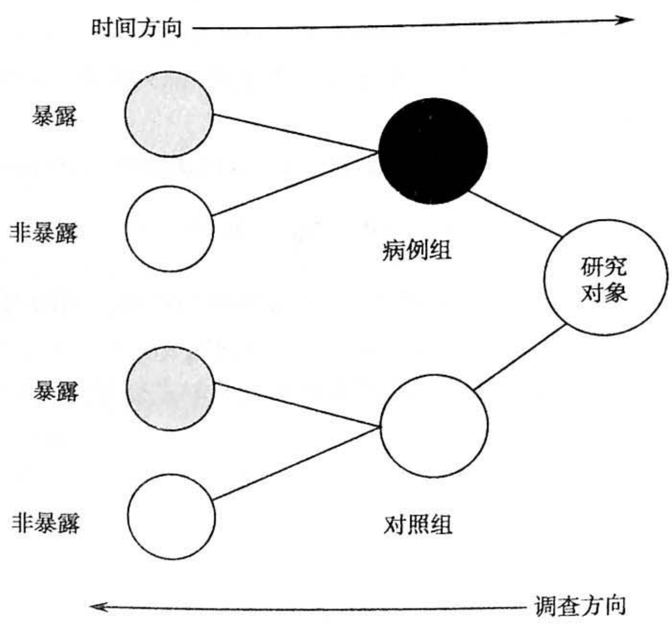


图 4-5　病例对照研究设计基本原理示意图

1. **成组病例对照研究**　在设计时对病例组和对照组人群在数量上没有严格的配比关系，对照组人群数量可等于、多于或少于病例组人数。

笔记

2. 配对病例对照研究　要求对照组在某些因素或特性上与病例组保持相同，形成匹配关系，而且数量上也要是配比关系，如1∶1或1∶2等。

（二）巢式病例对照研究

病例对照研究与队列研究作为探讨疾病病因等观察性研究的主要方法，已为国内外广泛使用。由于此两种方法各有其优势与不足，且两法正好优势互补，因此在实践过程中产生了这一扬长避短的新研究类型。

巢式病例对照研究（nested case-control study）是队列研究和病例对照研究的结合，其设计模式是：首先设计一项队列研究，收集基线资料，采集所研究生物学标志物的组织或体液标本储存备用，继之随访至能满足病例对照研究样本量的病例数为止。将这些病例作为病例组，按病例进入队列的时间与性别、年龄等配比条件，从同一队列中选择1个或数个非病例作对照，抽取病例与对照的基线资料并检测收集的标本，资料按病例对照研究的分析方法进行统计分析和因果推论。这种设计模式的主要优点是研究对象选择偏倚小，可以较好地避免回忆偏倚，研究和统计检验效率高，论证强度明显强于传统的病例对照研究。

三、实施方案

在提出开展病例对照研究的病因假设后，就需要制订出一套科学合理的实施方案。

（一）研究对象的选择

研究对象选择的原则有两个：第一，具有代表性：病例组应能代表目标人群中患该病的总体；对照组能代表目标人群中未患该病的总体；第二，具有可比性：病例组与对照组在年龄、性别、居住地、社会经济文化等主要特征方面应可比。

1. 病例组的选择　被选择的病例，诊断必须正确可靠，不能将诊断不明或误诊的病例作为"病例组"，否则会产生错误分类偏倚而造成低估疾病与暴露因素的关系；其次，被选择的病例，应具有暴露于调查（研究）因素的可能性，否则应予排除。例如，探讨口服避孕药物与某些疾病的关系时，对做过绝育术或因其他原因而忌用口服避孕药物者则不能选入；此外，应纳入新病例作为研究对象，以减少回忆偏倚的影响等。

关于病例组的来源，宜在同一地区不同水平的医院选用一个时期内符合要求的连续性病例，旨在一定的程度上防止选择性偏倚的影响。若条件许可，可选择社区总体人群中的全部病例（适于患病率低的疾病）或者从中随机抽样，其代表性更好。

2. 对照组的选择　被选择的"对照"，必须确实排除患有所研究的疾病，否则，也会出现错误分类；其次，"对照组"的研究对象也应具有暴露于被研究因素之可能性，并应与"病例组"同源（医院、社区等）。

3. 病例组与对照组的比较方式

（1）成组法：按和病例可比的原则，选择一定数量的对照。对照与病例的数量不需呈严格的比例关系。此法较配比法易于实施，但不易控制混杂因素。

（2）配对法：每一个病例选择一个或几个对照，使病例与对照配成对（pair 或 block），而对照在某些重要特征（如年龄、性别等）方面应与其相配的病例相同或基本相同。这些特征称之为配比因素（matching factor）。通过配对，可使病例组与对照组有可比性，较好地控制混杂因素。

病例与对照的比例，一般为1∶1，也可1∶2，但不超过1∶4。应注意：被研究的因素不能作为配比因素。配比的因素不应过多，否则容易发生"配比过度（overmatching）"，不仅影响结果的可靠性，而且浪费精力。

在研究中，可设多组对照，如既选医院的病人，又选亲属或邻居作为对照。这不仅扩大了对照的来源，减少偏倚，增强代表性；同时还可研究疾病与被研究因素在不同水平之间的关系或发现另外一些病因线索。

笔记

（二）样本量的估计

开展病例对照研究时，样本量的估算需要预先掌握4种参数，包括病例组和对照组各自研究对象对被研究因素的暴露率；预期优势比（OR）；容许误差的α值和β值。病例组和对照组样本含量相等时统计效率最高。估计样本含量的公式如下：

$$n=\frac{\left[Z_\alpha\sqrt{2\overline{P}(1-\overline{P})}+Z_\beta\sqrt{P_1(1-P_1)+P_0(1-P_0)}\right]^2}{(P_1-P_0)^2}$$

或简便公式：

$$n=\frac{(Z_\alpha+Z_\beta)^2\times 2\overline{P}(1-\overline{P})}{(P_1-P_0)^2}$$

n：样本含量；P_0：对照组暴露率；P_1：病例组暴露率；$\overline{P}=(P_1+P_0)/2$。其中，Z_α、Z_β可根据标准正态分布简表查出；P_1可根据P_0与OR推算，公式为：

$$P_1=ORP_0/(1-P_0+ORP_0)$$

（三）资料的收集

无论设计多么严谨，资料收集过程中若方法不恰当，将会产生无法纠正的系统误差，因而质量控制十分重要。

1. 资料来源 主要来源于设计良好的调查问卷，如果医院病例记录、疾病登记报告等能够满足研究所需，也可从中摘录，对调查表进行补充。

2. 收集方法 主要通过询问调查、查阅病历等方法收集资料，最常用的是访谈、信访及电话调查等，还可以通过查阅资料来收集。

3. 调查表设计原则 调查表的设计绝非易事，一张好的调查表的设计需要临床医学、流行病学、统计学、心理学和社会学的专家共同讨论拟定，并经反复修订和预调查后，最终形成后方可用于正式调查。调查表设计的基本原则包括：①调查条目的设置要全而精；②每个条目应定义明确；③调查条目应有具体的量化标准；④条目中的问题要通俗易懂，尽量口语化等。

四、资料的整理与分析

对原始资料进行整理、核查和计算机录入，在必要的数据转换后，首先要检验病例组与对照组是否具有均衡性，即在研究因素以外的其他主要特征方面的可比性如何？计数资料常用χ^2检验，计量资料常用t检验。如果组间差异无统计学意义，提示两组的可比性较好。然后对所研究的暴露因素进行逐项整理统计，计算OR值和其他指标；若存在混杂因素则应做分层分析，涉及多因素者则需进行多因素分析。在此基础上，对被研究因素和疾病的关系作结论。

（一）成组病例对照研究资料的结果分析

如果设计、资料收集是按成组设计进行的，结果分析同样应按成组设计进行。成组法资料整理的四格表见表4-6。

表4-6 成组病例对照研究结果分析用四格表

暴露	病例组	对照组	合计
有	a	b	$a+b$
无	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d=N$

1. χ^2检验 分析暴露与疾病之间是否有统计学关联。

基本公式：

$$\chi^2=\frac{(ad-bc)^2N}{(a+b)(c+d)(a+c)(b+d)}$$

笔记

2. 优势比（odds ratio, *OR*）　病例对照研究计算的是两组暴露率之间的比值，也称为比值比。*OR* 值的计算公式为：$OR=ad/bc$。

OR 值表示病例组暴露于某危险因素的比例是对照组暴露比例的多少倍。当 *OR* 值大于 1 时，说明暴露因素与该疾病呈正相关，疾病的危险度增加，*OR* 值越大，危险性越大；*OR* 值小于 1 时，说明研究因素与该疾病呈负相关，疾病的危险度减少，*OR* 值越小，保护作用越强；当 *OR* 值等于 1 或接近于 1 时，说明暴露因素与患病之间无联系。

（二）配对病例对照研究资料的结果分析

成组资料中的数字表示的是病例组和对照组的人数，而配对资料中的数字则表示的是对子数。对于 1∶1 的配对资料来说，表格中的数字表示 1 个病例和 1 个对照，若为 1∶2 的配对资料，则表示 1 个病例和 2 个对照，在分析资料时，以对子数为基础，不拆开进行分析，最为常用的 1∶1 配对的病例对照研究资料整理如表 4-7。

表 4-7　配对病例对照研究结果分析用四格表

对照	病例		合计
	有暴露史	无暴露史	
有暴露史	a	b	$a+b$
无暴露史	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d=N$

χ^2 检验（McNemar 检验）和 *OR* 值计算公式如下：

1. χ^2 检验

$$\chi^2=\frac{(|b-c|-1)^2}{b+c}$$

2. *OR* 值

$$OR=\frac{c}{b}$$

1∶2、1∶3 和 1∶4 的资料整理表和 χ^2 检验、*OR* 值计算公式与 1∶1 配对的病例对照研究不同，参见相关书籍。

五、应用范围

病例对照研究主要用于疾病危险因素的探索，但也可用于临床筛检、治疗效果评价等的研究。

1. **探索病因和危险因素**　从 20 世纪中叶开始已有大量有关疾病病因的病例对照研究。如吸烟与肺癌的关系、孕早期服用沙利度胺（反应停）与婴儿短肢畸形等很多经典的病例对照研究，都为探明相关疾病的病因起到了决定性的作用。

2. **评价筛检试验效果**　20 世纪 80 年代以来，很多学者应用病例对照研究方法对宫颈涂片检查宫颈癌、乳房照片检查乳腺癌、X 线胸透筛检肺癌等进行了科学评价。

3. **评价干预和治疗效果**　病例对照研究特别适合于发生率很低的某些疾病或事件研究，因为此时很难进行随机对照试验（RCT）。

4. **研究药物的不良反应**　当高度怀疑某种药物可能存在某些不良反应时，病例对照研究常是切实可行的方法，此时 RCT 等因伦理学限制而无法实施。

实例：在国家“十一五”传染病防治科技重大专项的支持下，第四军医大学课题研究组在甘肃省武威市开展了一项乙肝感染基线调查，结果发现，尽管当地婴儿的乙肝疫苗接种率已达 98% 以上，但 HBsAg 阳性母亲所生婴儿的乙肝感染率仍然达到 7.29%（18/247），原因何在？

笔记

为此，我们联合当地妇产科医生对婴儿乙肝感染的危险因素进行了一项病例对照研究，以2008—2010年在武威市3所医院妇产科分娩的HBsAg阳性母亲及其婴幼儿作为研究对象，收集这些母亲孕期、分娩和产后的医学事件；调查婴儿出生后乙肝疫苗接种时间、剂量等情况，检测母亲及其婴儿乙肝病毒感染标志物。以HBsAg和(或)HBV DNA阳性的婴儿为病例组，阴性者为对照组，应用单因素和多因素logistic回归模型分析母婴传播的危险因素。结果表明：母亲HBV DNA阳性（OR=4.83，95% CI：1.38~16.98，P=0.0140）、早产儿或低体重儿出生时推迟接种首针乙肝疫苗（OR=9.73，95% CI：1.78~53.21，P=0.0087）是该地区乙肝病毒母婴传播的主要危险因素。因此，课题组建议乙肝防治指南中应对该部分高危人群的乙肝疫苗接种进行具体指导。在循证医学证据表明乙肝疫苗接种对早产儿和低体重儿无毒副作用的情况下，所有婴儿均应执行出生24小时内接种乙肝疫苗的规定。该建议已为当地卫生部门所采纳，而且新生儿乙肝疫苗及时接种率也纳入到了当地医院医疗质量的考核指标中。该研究成果发表在了国际杂志*Vaccine*上。从中可看出病例对照研究在探索疾病发生的危险因素方面具有突出的实用价值和现实意义。

六、优 缺 点

1. 优点 所需样本量小，研究对象易找，工作量小，人力、物力也较少，因此易于进行，出结果快，可以对一种疾病的多种病因同时进行探讨。往往是罕见病病因研究的唯一设计模型。

2. 缺点 主要是容易受到回忆偏倚的影响，合理的对照选择又较困难，偏倚可能较大，论证强度不高；另外，病例对照研究无法计算发病率，只能推算出优势比。

（闫永平）

第六节 横断面研究

横断面研究（cross-sectional study）常归类于描述性研究设计，但其研究对象的选择、影响因素的调查及其结果分析较其他描述性研究（如病例报告等）更为严密和规范；而且通过对患者与非患者的分布特征和影响因素分析还可以进行一定程度的比较性研究。因此，横断面研究常常是进一步开展病例对照研究与队列研究等的前期基础，尤其是在疾病的患病率调查和人群、地区等分布特征研究中应用最为广泛。如：2010年开展的全国第五次结核病现况调查、2012年启动的全国精神障碍流行病学调查等，这些研究为我国相关疾病的防治和卫生决策制订提供了第一手资料。

一、概 述

（一）概念

横断面研究是指某一时点（或期间）内对某一特定人群中的疾病患病（或事件发生）状况及其影响因素（暴露）进行的调查分析。由于是在短时间内完成，如一天、一周或一个月，且调查的是患病频率，因此又称现况研究，或现患率研究（prevalence study）。

（二）特点

1. 在设计阶段一般不预设对照组，往往是先对全部研究对象进行调查后，在资料处理阶段，再根据患病与否或暴露状态进行分组比较。

2. 调查时间是某一特定时间点或时间段，对于患病率来讲，这个时间应该是越集中越准确。

3. 在确定因果联系时受到限制，大多仅能提供病因线索。

4. 在同一人群定期重复开展横断面调查可获得发病率、新发感染率、转归等资料。

笔记

二、设计模式

横断面研究主要是通过普查或抽样方式，研究目标人群疾病的患病率及其暴露状况，设计的基本原理如图 4-6 所示。

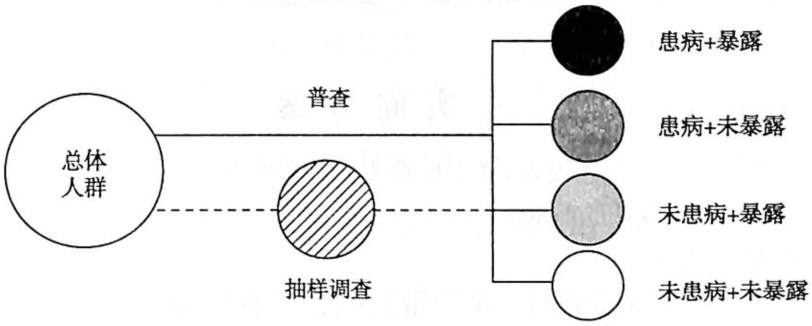


图 4-6　横断面研究设计原理示意图

进行横断面研究的方式分成两种：即普查与抽样调查。

1. 普查（census）　是指在特定时间对特定范围内的全体成员（总体）进行的全面调查。普查的目的主要是：①可以早期发现、早期诊断和早期治疗病人，如很多地区开展的妇女宫颈癌普查；②了解总体健康状况或某种疾病的患病率；③了解人体各类生理生化指标的正常参考值范围，如各国开展的儿童身高、体重等发育和营养状况的普查。该种方式调查全面，但相对费时、费力，质控较难。

2. 抽样调查（sampling survey）　简称抽查，是指在特定时间从特定范围内的全体成员（总体）中抽取具有代表性的部分成员进行的调查。

抽样的基本原则是随机抽样（random sampling）。在实际工作中常用的随机抽样方法有下列几种：

（1）单纯随机抽样（simple random sampling）：又称简单随机抽样，是指从总体中抽取了若干个体，构成一个样本。抽样过程中不附加任何限定条件，在抽样前未进行分层或其他方式处理，保证总体中的每个个体被抽到的机会相等。抽样的基本方法是使用随机数字表法或计算机随机数字方法。单纯随机抽样是比较简单而可靠的随机化方法，但在总体与样本都比较大时，编制名单、号码、卡片或抽样都会耗费大量的人力、物力。同时入选个体若散在分布于一个地域广袤的地区内，调查难度将大大增加。因此调查中很少单独使用简单随机抽样法，往往将其作为其他抽样方法的基础。

（2）整群抽样（cluster sampling）：是以多个个体组成的群组为单位而进行的随机抽样，对被抽中单位内的每个成员都进行调查。这种抽样和调查都比较方便，特别适合于大范围的横断面调查研究。其优点是便于组织、节约人力和物力，在实际工作中易被群众接受；缺点是抽样误差较大，分析工作量也较大。

（3）分层抽样（stratified sampling）：为了保证调查对象的同质性，可以按主要影响因素如年龄、性别和职业等先分成若干层，大型断面调查则可按行政单位或地区分层。各层内再作简单随机抽样。该法的优点是分层后，层内各单元的特征比较齐同，变异较小，精确度提高，从而有利于节省样本量。

（4）系统抽样（systematic sampling）：系统抽样是指按照一定的顺序，每隔若干单位机械地抽取一个单位的抽样方法，又称机械抽样。优点是简便易行、样本的观察单位在总体中分布均匀，抽样的代表性较好，抽样误差接近甚至略小于单纯随机抽样法。缺点是如果总体各单元的排列顺序有周期性规律，抽取的样本可能存在偏倚。

笔记

（5）多阶段抽样（multi-stage sampling）：是指从总体中先抽取范围较大的单元，一般称为一级抽样单元（例如省和市等），再从抽中的一级单元中抽取范围较小的二级单元（如县和乡等）……依此类推，最后抽取其中范围更小的单元（如社区、行政村）作为调查对象。

在实际的调查工作中，普查与抽查往往是相对的，如整群抽样时，在被抽中的基层单位内实际上是进行了普查，但在总体看来却是抽查。在规模较大的横断面调查中，各种抽样方法常常结合运用。

三、实施方案

横断面研究的规模一般都比较大，涉及的调查人员和调查对象也较多，因此有一个良好的设计实施方案是保证研究成功的前提。

（一）研究对象的选择

基于研究目的，对调查对象的特点和范围进行界定，同时还要结合实际情况，考虑在目标人群中开展普查或抽样调查的可行性。在设计时可以将研究对象规定为某个社区或乡镇内的全体居民或其中的一部分，如儿童，即选择该社区内≤14岁者。

（二）样本量的估计

样本的大小与抽样的方法有关。简单随机抽样的样本量的大小与预期患病率和调查的精确度要求有关。预期的患病率（P）或感染率高，样本量则可以小一些；要求的精确度高，则样本量大。精确度又与容许误差（d）和显著性水平（α）有关。其样本量的大小可以用下列公式进行计算：

$$n=K\cdot Q/P$$

式中 n 为样本量，P 为预期患病率，$Q=1-P$，K 为系数。在 $\alpha=0.05$ 水平上，当容许误差 d 为 $0.1P$ 时，$K=400$；当容许误差为 $0.15P$ 时，$K=178$；当容许误差为 $0.20P$ 时，$K=100$。

（三）资料的收集

所收集的资料信息因研究目的的不同而有所差异，通常主要包括：个人基本情况（如年龄、性别、文化程度、籍贯、职业及所在单位等）、人口学资料、生活习惯、环境资料等。相关资料一般可从临床和实验室检查、调查询问和常规资料记录中获得。

四、资料的整理与分析

在进行资料的整理和分析时，要根据不同的研究目的，选用不同的指标。除暴露因素外，对疾病或事件的频率描述，常见的有：患病率、抗体阳性率、抗原或病毒携带率、疾病诊治率、疫苗接种率、伤残率等。

患病率（prevalence），可按调查时间的长短分成时点患病率（point prevalence）与期间患病率（period prevalence）。时点患病率的时间常指某一天（如12月31日）或更短时间；期间患病率则可以比较长：周、月甚至更长些。调查的病例数（即计算患病率时的分子），应包括调查时点或期间内、新发生的、正在患病的，以及在此期间内发病后病故的全部病例。

不同地区间患病率进行比较时，常采用率的标化方法。而相关影响因素的分析可采用单因素和多因素统计分析模型来完成。

五、应用范围

1. **描述疾病的分布** 通过计算和比较所获得的患病率与感染率等指标，描述目标人群疾病的年龄、性别、种族、地区等的分布状况，从而为制定防治对策提供参考依据。

2. **提出病因线索** 可以从疾病与研究对象特征以及环境因素的联系中研究疾病分布的不同影响因素，从而提出病因线索。

笔记

3. 确定高危人群　通过筛查，可以检出患者、可疑患者和病原携带者，从而确定高危人群，达到早发现、早诊断和早治疗的目的，提高疾病诊治效果。

4. 评价防治效果　在疾病监测、预防接种的实施过程中，通过不同阶段开展重复的横断面调查，比较患病率、感染率等相关指标的差异，评价防治方案的确切效果。

5. 医疗卫生服务的需求与质量评价　通过横断面调查可以评价社区卫生服务的近期和远期效果。近期效果如体检率、家庭和个人健康档案建档率、疫苗接种率、居民卫生知识知晓率等。远期效果如发病率下降、死亡率下降、人均期望寿命延长和医疗费用下降等。

六、优缺点

1. 优点　容易实施，科学性较强，研究对象代表性好，一次研究可观察多种疾病（事件）的患病状况及多种相关的可能影响因素。

2. 缺点　一次横断面调查难以确定暴露与疾病之间的因果关系，尤其是在开展大规模调查时，需投入很多人力、物力。

（闫永平）

第七节　病例分析

病例分析（case analysis）是临床医生最为熟悉和常用的一种研究方法。可为医疗实践提供最基本的信息，发现临床试验中容易忽略的问题，产生新的研究思路，更好地服务于患者。病例分析往往反映研究者临床经验的积累水平、对当前学科发展方向的把握以及对相关领域研究结果的关注度。

一、概　述

（一）概念

病例分析是对现有的临床资料进行归纳、分析并得出结论，或对某些临床新出现的疾病病因或表现特征进行描述、分析、总结的一类研究。主要包括个案病例报告和系列病例分析等，成为临床医生日常工作中使用最广泛，也最易掌握的一种描述性研究方法。

（二）特点

1. 以观察法为主要研究手段，主要通过观察、收集和分析相关病例数据，归纳和总结研究对象的基本特征。

2. 由于不设立对照组，仅能提供因素与事件之间的因果联系线索，但可为后续分析性研究打下基础。

二、设计模式

（一）个案病例报告

个案病例报告（case report）是对单个或少数几个病例的个人基本信息、临床和流行病学特征的描述、分析和总结。尤其在罕见和新发疾病的报道中最为有用，常常为发现新病种或药物副作用等提供第一手资料。

（二）系列病例分析

系列病例分析（case series）是对一系列或一组病例的人口学特征、临床和流行病学特征的描述、分析和总结。在设计时一般采用以下 2 种模式：

1. 连续系列病例分析（consecutive case series）　就是将研究期间内所有符合条件的病例均纳入分析。

笔记

2. **非连续系列病例分析**（nonconsecutive case series） 就是将研究期间内部分符合条件的病例纳入分析。一般采用随机抽样方法选择这部分病例，作为总体病例的代表。

三、实施方案

实施病例分析时，首先是要根据临床观察和资料报道提出拟分析的问题，进一步查阅、了解与该问题相关的文献，阅读相关的病历记录，制订调查和随访内容，然后确定病例范围和研究期限，明确纳入和排除标准，以保证结论的可靠性。

（一）研究对象的选择

病例系列或一组病例往往选择接受同样的治疗方案者，以便分析疗效和药物副作用等；或者选择疾病诊断相同的病例用以分析不同方案的疗效及其影响因素等。

（二）样本量的估计

系列病例分析样本量原则上是越多越有说服力。但要注意的是如果样本量过大时，要注意控制好系统误差。

（三）资料的收集

需要收集详细的人口学特征（如年龄、性别、种族等）、主诉、症状、体征、体格和实验室检查、既往史、家族史、诊断、治疗、药物反应和随访等重要信息。除病历和随访记录外，还可以设计简要的调查表收集资料，对记录不清楚、项目不完整者进行补充调查。

四、资料的整理与分析

一方面，可以观察干预措施对疾病的治疗效果和药物副作用等。例如，接受化疗的恶性肿瘤病人，观察其化疗的不良反应或并发症的发生率、发生的时间、严重程度和恢复的时间等；另一方面，还可从已获得结果中分析和总结出可能的病因或干预效果。例如在肺癌患者中，调查吸烟的历史。从高血压患者并发症分析中评价降压药的作用和效果等。另外，还可基于病例分析的具体目的进行相应地分析和总结，重点是探讨研究中的新发现以及对临床实践的应用价值。

结果分析中，计数资料常采用率表示，计量资料则用均数 ± 标准差表示。并以 95% 可信区间来表示结果的精度。

五、应用范围

病例分析在临床上应用最为广泛，几乎涉及临床的方方面面。如描述罕见或新发现疾病的临床特征、诊治方法和预后，描述新的手术方式和医疗革新，描述危及患者生命、罕发的药物副作用，报告医疗事故、差错和经验教训，总结临床治疗和护理经验等。

个案病例报告主要描述一种新的疾病或综合征。一些疾病就是通过病例报告而为人所知，如多发性骨髓瘤的发现。当一定数量的个案汇集到一篇文章中，就被称为系列病例分析，可及早发现疾病流行的端倪，了解药物的不良反应，如 1981 年洛杉矶发现多个艾滋病患者。倘若个案病例还可以简单地归为偶然事件，但系列病例发生时就需要考虑深层次的原因。如，乳房硅胶植入与系统性红斑狼疮的可能联系。但这只能是线索，如果不做进一步检验和验证，往往会出现误判。

实例：2003 年 10 月《中华结核和呼吸杂志》报道了世界首例严重急性呼吸综合征（SARS）的病例报告，引起广泛关注。作者在该文中详细描述了该例病人的一般特征和诊治过程。患者为男性，43 岁，既往体健，广东省佛山市人，于 2002 年 11 月 16 日开始出现不明原因的持续发热（>39℃）伴腹泻、咳嗽，无痰，无鼻塞、流涕，无皮疹和肌肉酸痛。在当地一诊所查血：白细胞计数（WBC）为 3.4×10^9/L；胸部 X 线：双肺纹理明显增强，以右肺明显；拟诊"支气管肺炎"，使用二、三代头孢菌素和三唑巴坦加用哌拉西林等治疗无效。11 月 25 日转至佛山市第一人民医院。

笔记

胸部X线示双肺弥漫浸润。经过吸氧、静脉注射头孢他啶和环丙沙星等治疗，至11月28日，体温降至正常范围，腹泻控制，但呼吸困难等症状加重，动脉血气分析提示严重低氧血症，拟诊"急性呼吸窘迫综合征（ARDS），Ⅰ型呼吸衰竭"转入ICU。11月29日出现肝功能不全等多脏器功能不全综合征（MODS），胸部X线示双肺渗出明显，经过气管插管、机械通气、抗病毒、抗生素和糖皮质激素加用分子吸附再循环系统（MARS）等治疗19天，病情逐渐好转，2003年1月8日治愈出院，但双肺阴影仍未完全消退，直至出院第49天双肺阴影完全消退，肺纹理转为清晰。该病例报告为国内外SARS防治研究提供了宝贵的第一手资料。其实在整个SARS发现和防治过程中，从首例的个案报道、系列病例分布特征、实验室检查和治疗经过等的描述，到分离出新型冠状病毒，最终对疫苗的研制和激素等的综合治疗，可以清楚地看到，一种未知原因的疾病的研究，常常是以个案报道及系列病例分析为起点的，因此，病例分析仍然是临床研究不可或缺的设计方案。但是它的偏倚也是无法克服的，如该首例SARS病例到底有无传染性？病因是什么？诊断标准如何制定？治愈究竟是哪种或哪几种药物和治疗手段发挥了作用呢？抑或是自愈？等等。这些问题都需要进一步采用分析性和实验性研究方法来解决。

六、优　缺　点

1. 优点　研究容易实施，节省人力、物力，短期易出结果，常作为很多临床分析性研究和实验性研究的基础。在个别情况下，设对照组可能没有必要，系列病例报告就能很好说明问题。例如，如果调查出生1个月内婴儿死亡的原因时发现有很多婴幼儿是由于头部跌落在地板上所致，此时再设置对照组开展分析和实验性研究就显得没有必要，而且与伦理学相悖。再比如，使用链霉素治疗结核性脑膜炎，不治疗是致命的，那么存活率的普遍提高就是显著的效果。

2. 缺点　没有对照组，大多情况下研究结果的说服力不强。不能控制选择偏倚和混杂因素对结果的影响，研究结果的论证强度弱，重复性较差。

（闫永平）

笔记

第五章　医学研究证据的检索与收集

哈佛大学医学院院长 Sydney Burwell 博士曾说过："医学生在学校接受的知识，10 年后其中一半可能是错误的，而可悲的是没有人能预测哪一半是错误的。"因此，为了不被过时、错误的知识所误导，需要针对临床工作中遇到的问题，不断查寻新的文献资料，掌握学科发展的历史、现状和把握未来发展方向，更新知识结构。而繁忙的临床医师如何在众多的医学刊物和数据库中快速、有效地查寻相关领域的最佳信息，无疑是一种挑战。

第一节　循证医学的证据资源

临床医生面临临床问题时常查寻的信息资源包括教科书、专著、中国生物医学文献数据库、MEDLINE 等或咨询专家。然而，全世界有为数众多的医学文献数据库，每个专业或亚专业又有其独特的数据库资源或检索工具，同时并不是每个数据库提供的信息资源都是真实、可靠且能方便、快捷地获取的。

一、循证医学的证据资源及其发展

循证医学的证据资源经历了漫长的发展过程。20 世纪 90 年代前，临床医生主要依靠手工检索工具和光盘先获取题录，再查阅全文。这种方式耗时且不一定能获得最新证据信息。随着计算机和网络技术发展，信息的储存、传输和利用进入了一个崭新的阶段。人们逐步摆脱对耗时、费力、检索效果差的手工检索依赖，转而采用快速、高效的联机检索，既克服了时空障碍，又极大提高了获得最新信息的能力。大存贮量的光盘和联机检索极大地降低了计算机检索的费用，使及时、准确、全面利用人类研究成果，成为可能。在此基础上，一些信息专家、方法学家和临床专家们开始考虑如何从浩如烟海的信息资源中去粗取精、去伪存真，为繁忙的临床医生提取真实、可靠且有临床实用价值的信息，这些资源包括 1991 年创刊的美国内科医师学会杂志俱乐部（ACP Journal Club）、1993 年 Iain Chalmers 等创立 Cochrane 协作网和建立 Cochrane 图书馆、1999 年 BMJ 推出的 Clinical Evidence，越来越多的循证医学信息资源可供临床医生选择。这些资源或从方法学质量和临床价值方面严格评估原始临床研究，或采用系统评价和 meta 分析方法整合针对同一临床问题的高质量原始研究，或针对临床主题总结治疗某一种疾病的所有证据，帮助临床医生明确哪些干预措施有效、哪些无效甚至有害，极大方便了临床医生开展循证临床实践。随后，各大医学信息数据库提供商相继推出 PIER（Physician's Information Education and Resources）、DynaMed、UpToDate 等数据库，集多种高质量资源于一体，在总结证据基础上，结合专家经验、患者价值观给出推荐意见，进而基于证据的质量给出推荐强度。至此，临床医生不再需要自己花费大量时间从 PubMed、Embase 等原始文献数据库中去检索、获取全文、评价和总结临床研究证据。这类资源的不断丰富和发展，使实践循证医学成为可能。

国内目前还没有类似的中文数据资源，使用这些英文的资源，国内医生将面临语言、医疗环境差异和费用等难题。

二、循证医学证据资源分类

Haynes 等于 2001 年、2006 年和 2009 年分别提出了循证医学资源的"4S"、"5S" 和 "6S" 模型，细化了证据资源的分类，为不同用户提供更新、更实用且经过筛选的循证医学证据。"6S"

笔记

模型将信息资源分为6类，即系统类(system)、集成类(summaries)、系统评价摘要类(synopses of syntheses)、系统评价类(syntheses)、原始研究摘要类(synopses of studies)和原始研究类(studies)。

1. 系统类 即计算机决策支持系统(computerized decision support system, CDSS)，是指针对某个临床问题，概况总结所有相关和重要的研究证据，并通过电子病例系统与特定患者的情况自动联系起来，为医生提供决策信息。现有的数据库尚不能达到如此高智能化程度。即使有，证据系统本身也不能直接告诉医生如何治疗患者，需要临床医生根据经验和专业知识将证据与患者的具体情况和期望整合进行决策。

2. 集成类 针对临床问题总结研究证据，但不与具体病案链接的循证医学数据库资源，包括针对具体临床疾病的临床路径、不断更新的网络版教材以及循证临床实践指南等，这类资源有：Clinical Evidence(http://www.clinicalevidence.com)、由美国医师学院提供的PIER(Physician's Information and Education Resource, http://pier.acponline.org/index.html)、Dynamed(www.ebscohost.com/dynamed)、UpToDate(http://www.uptodate.com)、ACP Medicine(www.acpmedicine.com)、ACS Surgery(www.acssurgery.com)、Harrison's Practice(www.harrisonspractice.com/practice/ub)、First Consult(www.mdconsult.com/php/142221516-2/homepage)、eTherapeutics(www.e-therapeutics.ca)、National Guidelines Clearinghouse(www.guidelines.gov)、Registered Nurses' Association of Ontario(www.rnao.org)等。

3. 系统评价摘要类 精心编辑的结构式摘要，通常仅1页，由方法学家和临床专家从发表的系统评价中筛选出的临床相关且有重要临床价值的高质量证据，主要发表在美国内科医师学会杂志俱乐部(American College of Physician Journal Club, ACP Journal Club, http://www.acpjc.org)、*Evidence-Based Mental Health*(http://ebmh.bmj.com)、*Evidence-Based Nursing*(http://ebn.bmj.com)和系统评价摘要库(DARE, the *Database of Abstracts of Reviews of Evidence*, www.crd.york.ac.uk/crdweb)等。

4. 系统评价类 系统评价是针对某一具体临床问题(如疾病的病因、诊断、治疗、预后)，系统、全面收集全世界所有已发表或未发表的临床研究，严格评价纳入文献的偏倚风险，筛选出符合质量标准的文献，进行定性或定量合成(meta analysis，荟萃分析)，得出可靠的综合结论。相对于单个原始临床研究，系统评价对精力、时间有限的临床医生来说更实用。系统评价分为Cochrane系统评价和非Cochrane系统评价，前者由Cochrane协作网(http://www.cochrane.org/index.htm)的作者制作并发表在Cochrane图书馆，占全世界系统评价的30%~40%，后者发表在杂志上。相关数据库EvidenceUpdates(http://plus.mcmaster.ca/EvidenceUpdates)。

5. 原始研究摘要类 由方法学家和临床专家对主要医学期刊上发表的原始研究从方法学和临床重要性两方面评价，筛选出高质量论著以结构摘要的形式再次出版，并附专家推荐意见。如美国内科医师学院杂志俱乐部(American College of Physician Journal Club, ACP Journal Club, http://www.acpjc.org)和InfoPOEMs(http://www.infopoems.com/)。

6. 原始研究类 发表在杂志和综合文献数据库、未经专家评估的文献资料如MEDLINE、Embase、中文数据库(CBM、CNKI、CMCC、VIP)等。临床医生在检索和应用此类文献时，需要自己进行评估研究结果的真实性、临床重要性和适用性后方可应用，否则可能误导。

三、证据资源的入选标准

针对某一临床问题，要快捷、高效检索到相关信息，正确选择信息资源或数据库十分重要。如何选择信息资源呢？McKibbon等列出了4条标准(表5-1)。

1. 循证方法的严谨性 循证医学信息资源需要针对临床问题提供具有代表性的高质量证据。在总结证据资源时，应精心构建临床问题、系统全面检索相关证据、严格评价单个研究的

笔记

表 5-1　选择或评估信息资源的标准

选择标准	标准内容描述
1. 循证方法的严谨性	✓ 推论是否严格遵循证据
	✓ 提出推荐意见时是否给出支持其结论的证据强度
	✓ 是否为读者提供了证据链接以方便阅读
2. 内容的全面性和特异性	✓ 是否充分覆盖了我的专业领域或内容范围
	✓ 是否覆盖了我提出的问题类型(治疗、诊断、预后、病因或不良反应)
	✓ 是否针对我临床实践的具体专业领域
3. 易用性	✓ 能否快速、始终如一地提供我需要的信息
4. 可及性	✓ 是否在我需要使用的任何场所均能方便获取
	✓ 是否能支付其费用

真实性、恰当总结和合成结果。循证医学信息资源在提出推荐意见时要充分应用已有的系统评价或自己制作系统评价，针对患者重要的结局指标(patient-important outcomes)，为不同干预措施的疗效和安全性提供最佳估计，并在充分考虑患者的价值观和选择基础上，采用恰当的分级系统对推荐意见进行分级。

2. 内容的全面性和特异性　理想的循证医学信息资源应该为临床实践中可能遇到的所有问题提供相关证据。但针对某一专业领域的信息资源可能会更有效地帮助查寻需要的证据。如要紧跟某些临床专业的最新进展，可查寻某些循证的摘要如 Evidence-based Cardiovascular Medicine，Evidence-based Mental Health 和 Evidence-based Oncology 等。

某些信息资源是专门针对某一类临床问题，如 Clinical Evidence 和 Cochrane 系统评价数据库目前仅涉及治疗性临床问题，且后者仅包括临床对照试验的系统评价。

3. 易用性　某些信息资源查寻起来方便快捷，如美国内科医师学会杂志俱乐部(ACP Journal Club)为小型数据库，收集了内科领域约 140 种杂志上发表的与临床最相关的高质量研究的摘要，其优秀的检索引擎确保读者能很容易查寻到任何该领域的信息，包括疾病病因、诊断、治疗和预后等。

MEDLINE 是检索世界生物医学文献资源最主要的数据库之一，MEDLINE 的检索可采用联机检索、光盘检索，近年来由于计算机国际互联网(INTERNET)的应用，90% 的国际联机检索系统都已进入 INTERNET。通过隶属美国国立医学图书馆的国家生物技术中心开发的 INTERNET 免费 PubMed(http://www.ncbi.nlm.nih.gov/pubmed/)网络检索系统，可检索 MEDLINE 数据库及 Pre-Medline 数据库。数据库容量庞大，2008 年初已包含约 1700 万篇文献，每年以 70 万条速度增加，其检索速度较慢。PubMed 是检索 MEDLINE 较容易的途径，其专门为临床医生设计的"Clinical Queries"检索方式可将检索结果最大限度地局限于与临床决策相关的文献资料。

Cochrane 系统评价尽管容易检索，但每一篇系统评价包括内容较多、统计分析结果较复杂、报告质量各异，临床医生需要花很多时间阅读和理解方可应用于临床。

4. 可及性　最可靠和最有效的信息资源往往需要付费且价格比较昂贵。临床医生通常利用所在医院或医学院校图书馆免费检索网上资源，个人大多不会订阅昂贵的期刊杂志或数据库。目前大多可及信息资源是免费的，如中低收入的发展中国家的临床医生可通过学术机构的网络进入世界卫生组织的健康网络研究启动项目(Health InterNetwork Access to Research Initiative，HINARI)获取免费信息资源，对于全球免费开放的资源有：PubMed，Canadian Medical Association Journal(http://www.cmaj.ca)和大多数 BioMed Central 的杂志(http://www.biomedcentral.com)，一些杂志可在其发表 6~12 个月后全部免费检索，甚至可免费检索最新一期的部分文献，如 BMJ，JAMA 和 Mayo Clinic Proceedings 等。

笔记

四、常用循证医学证据资源

根据"6S"模型,下面简单介绍常用的各类数据库资源。由于证据系统是理想化、高智能的系统,目前尚无循证医学数据库能达到此水平,因此直接从证据集成开始介绍。

(一)证据集成

1. Clinical Evidence Clinical Evidence(http://www.clinicalevidence.com/ceweb/conditions/index.jsp)由BMJ出版,是世界上最具权威性的医学数据库之一,以内科治疗为主,涉及600多种疾病的3250多种治疗方法,每年更新一次并在不断拓展新的题目和领域如疾病诊断。该数据库针对每种疾病,严格评估每种治疗方法的疗效和安全性,告诉读者哪些治疗方法有益、哪些可能有益、哪些利弊相当、哪些不可能有益、哪些可能无益甚至有害和哪些疗效不确定等。此外,针对每个主题采取列表的方式列出最新发表的证据,及时更新内容。

Clinical Evidence方便易用,但除其成员和中低收入的发展中国家外,均需付费。北京大学的循证医学中心曾将其15版翻译为中文《临床证据》于2008年出版,方便国内读者,但存在及时更新问题。

2. PIER PIER(Physicians' Information and Education Resource)(http://pier.acponline.org/index.html)是美国内科医师学会的产品。主要包括:疾病、筛查与预防、补充/替代医学、伦理和法律问题、流程、质量测量方法和药物信息。其优点是采用多层次结构指导临床医生应用研究证据,所有问题均采用同样结构,所有推荐意见均与研究证据紧密相连。基于严格的循证医学方法,从精心构建问题、全面收集所有干预措施和以病人为中心的结局指标、评估单个研究的质量、采用高质量的分级系统、到充分考虑患者的价值观和选择等,最终形成PIER推荐意见。PIER主要涉及内科和初级保健方面的治疗问题,覆盖面有限。

PIER虽方便易用,但需要付费,PDA版本收费标准约100美元。ACP注册成员可免费。

3. UpToDate UpToDate(http://www.uptodate.com/index.asp)是一电子信息资源,可从网上、PC机和PDA获取,由于其使用方便、覆盖面广和根据疾病分类收集信息,深受全科医师、专科医师和家庭医师青睐。类似于PIER,UpToDate为临床医生提供推荐意见,方法严谨,提出问题结构统一、循证医学文献检索较为全面,最后采用GRADE分级评价证据质量和提出推荐意见。UpToDate明确承认患者价值观和选择权在临床决策中的重要性。

UpToDate覆盖了14个医学专业的7700个临床主题,包括约80 000页正文、图片,并与MEDLINE摘要、260 000条参考文献和一个药物数据库链接,每4个月更新一次。

UpToDate使用方便,但缺乏规范检索,需要付费。个人第1年订购费为450美元,以后为350美元,图书馆订购为10 000美元。

4. 临床实践指南 临床实践指南(clinical practice guideline,CPG)是证据强度很高的信息资源,有助于指导临床决策。其中,美国国家指南数据库(US National Guidelines Clearinghouse Database,NGC,http://www.guideline.gov/)收集了美国和全世界数千个指南并提供了2488个指南的摘要(截至2012年12月),涉及所有主题,指南制订按照循证医学原则和方法。NGC检索简单,同时比较针对同一主题的多个指南。

此外,还可在英国国立卫生图书馆(UK National Library for Health,http://libraries.nelh.nhs.uk/guidelinesFinder/default.asp?page=INTER)和安大略医学会网站(Ontario Medical Association,http://www.gacguidelines.ca/)查寻到指南,后者指南是经过严格评估的。

(二)系统评价摘要

1. ACP Journal Club 包括ACP Journal Club(http://www.acpjc.org/)、Evidence-Based Medicine杂志和系列以ACP Journal Club为模版的杂志,以纸质或网络版发行。ACP Journal Club先由工

笔记

作人员从130种杂志中筛选方法学严格、涉及临床问题、报告了重要临床结局指标的高质量原始研究和系统评价，再让临床医生从中选择对临床有重要价值和影响的文献，以结构摘要形式进行总结，并由一名临床专家评估文献的方法和提出临床应用的建议。比例大致为150:1，即150篇文献中可筛选出1篇合格者。

ACP Journal Club由美国内科医师学会出版，主要针对内科及其亚专业，也包括少量其他专业如儿科，月刊，需付费，有助于临床医生了解内科领域新进展。

2. Database of Abstracts of Reviews of Evidence(DARE) DARE(http://www.crd.youk.ac.uk/crdweb)是唯一经过质量评价的免费系统评价摘要库，由英国约克大学生产。每篇摘要包括对系统评价的总结和整体质量的严格评价，对无时间评价或不能获取系统评价全文的临床医生非常有用。DARE涉及主题广泛，覆盖干预措施疗效和数千篇针对诊断试验、公共卫生、健康促进、药学、手术、心理学、卫生体系和服务的系统评价摘要。DARE既是独立信息资源，也包含在Cochrane图书馆中，每年收录约600篇，检索简单方便。

（三）系统评价

1. Cochrane系统评价数据库 Cochrane系统评价数据库(Cochrane Database of Systematic Reviews, CDSR)发表在Cochrane图书馆(http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME), Cochrane系统评价是Cochrane协作网的评价员按照统一工作手册(Cochrane Handbook for Systematic Reviews of Interventions)，在相应Cochrane评价小组编辑部的指导和帮助下所完成的系统评价。由于Cochrane协作网有严密的组织管理和质量控制系统，严格遵循Cochrane系统评价者指导手册，采用固定格式和内容，统一的系统评价软件(RevMan)录入和分析数据、撰写系统评价计划书和报告，发表后根据新的研究定期更新，有完善的反馈和修改机制，因此Cochrane系统评价的质量比收录在Medline和其他数据库的非Cochrane系统评价质量更高。目前主要针对疾病防治、康复疗效和安全性的随机对照试验进行评价，诊断性试验的系统评价刚起步。2012年第7期的Cochrane系统评价数据库包括5131篇全文系统评价和2235篇研究方案。Cochrane系统评价可从Ovid、PubMed、光盘和Wiley网站获取。CDSR检索方便，摘要免费，但全文需要付费。

2. EvidenceUpdates EvidenceUpdates(http://plus.mcmaster.ca/EvidenceUpdates)是主要的且经过评价的系统评价单一数据源，由BMJ集团免费提供。EvidenceUpdates包括所有Cochrane系统评价、120多种临床杂志发表的系统评价及由美国AHRQ组织、加拿大药品和卫生技术机构委托完成的系统评价、英国NHS卫生技术评估等。

3. 其他资源 除上述资源外，也可从PubMed和Embase数据库查寻系统评价，但需要检索者自己评价质量。

（四）原始研究摘要

1. ACP Journal Club 由方法学家和临床专家对主要医学期刊上发表的原始研究从方法学和临床重要性两方面评价，筛选出高质量论著以结构摘要的形式再次出版，并附专家推荐意见(形同系统评价摘要)。

2. InfoPOEMs InfoPOEMs(http://www.infopoems.com/)类似于ACP Journal Club，有助于了解临床新进展，重点针对家庭医学。临床人员从100多种杂志中筛选出与家庭医生常见问题相关的文献进行评估和总结。

3. Bandolie Bandolie(http://www.jr2.ox.ac.uk/bandolier/)是为英国国立卫生服务中心提供的证据，选择的主题涉及各临床专业，评估的证据包括评论和推荐意见。

另外，美国医学研究院(New York Academy of Medicine, http://www.ebmny.org/jour-nal.htm)网站列出了包括ACP Journal Club在内的经过评估的信息资源库，可供链接查寻。

笔记

（五）原始研究

原始研究文献数量最为庞大，通常只有在上述四种数据库资源中未能检索到需要的文献时才检索原始研究数据库。

1. MEDLINE 及 PubMed Clinical Queries MEDLINE 是卫生研究和医疗实践的首选数据库。检索 MEDLINE 的途径很多，但多数通过 Ovid 和 PubMed 途径。

通过 PubMed（http://www.ncbi.nlm.nih.gov/PubMed/）可检索整个 MEDLINE 数据库，其 Clinical Queries（http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml）检索可直接获得与临床应用相关的文献资料。

2. Embase Embase（http://www.elsevier.com/wps/find/bibliographicdatabasedescription.cws_home/523328/description#description）是欧洲的大型医学文献数据库，与 MEDLINE 类似，但重点在药物和卫生领域。由于 Embase 数据库检索费用高，可及性差，临床医生少有采用。Embase 中约 70% 的条目未包括在 MEDLINE。

3. Cochrane 临床对照试验中心注册库 Cochrane 临床对照试验中心注册库（Cochrane Central Register of Controlled Trials，CENTRAL）是随机对照试验和半随机对照试验的数据库，该数据库由 Cochrane 协助网组织、协调和编制，采用计算机和手工检索相结合的方法，对期刊、会议论文集、MEDLINE 和 Embase 及其他文献数据库收录的刊物进行检索，确定其中的随机对照试验和半随机对照试验，为进行系统评价提供系统、全面和准确的原始研究文献库。

4. 中国生物医学文献数据库 中国生物医学文献数据库（CBM）是中国医学科学院医学信息研究所开发研制的综合性医学文献数据库。该数据库收录了 1978 年以来 1600 多种中国生物医学期刊及汇编、会议论文的文献题录，收录范围涉及基础医学、临床医学、预防医学、药学、中医学、中药学等生物医学的各个领域。该数据库的研制兼顾了与 MEDLINE 光盘检索系统的兼容性，其检索过程中使用的运算符号及功能与 MEDLINE 光盘相似。检索系统具有主题词表、中英文主题词轮排表、分类表、期刊表、索引词表、作者表等多种词表辅助检索功能，检索入口多，检索功能完备。

5. 中文生物医学期刊数据库 中文生物医学期刊数据库（Chinese Medical Current Contents，CMCC）是解放军医学图书馆研制开发的中文生物医学文献书目型数据库，也是目前检索国内生物医学文献最常用的光盘数据库之一。CMCC 数据库收录了 1994 年以来国内正式出版发行的生物医学期刊和一些自办发行的生物医学刊物 1400 余种的文献题录和文摘。涉及的主要学科领域有基础医学、临床医学、预防医学、药学、医学生物学、中医学、中药学、医院管理及医学信息等生物医学的各个领域。CMCC 数据库的数据与 CBM 的数据从 1994 年后大部分是相同的，因此检索 1994 年后的中文医学文献选用其中之一即可。但由于 CMCC 的数据更新周期为 2 周，因此检索最新报道的生物医学文献时可选用。

6. 中国期刊全文数据库 中国期刊全文数据库是中国知识基础设施工程（China National Knowledge Infrastructure，CNKI，http://www.cnki.net/index.htm）中最重要的数据库。CNKI 是目前世界上最大的连续动态更新的中国期刊全文数据库，收录了自 1994 年以来（部分回溯至 1979 和创刊年）国内 8200 多种重要期刊，内容覆盖自然科学、工程技术、农业、哲学、医学、人文社会科学等各个领域，全文文献总量 2200 多万篇。本数据库集题录、文摘、全文文献信息于一体，实现一站式文献信息检索；具有知识分类导航功能、众多检索入口等功能。

7. 中文科技期刊数据库 中文科技期刊数据库（VIP）由重庆维普资讯有限公司开发，是一种综合性文献数据库，收录了 1989 年以来 8000 余种中文科技期刊，涵盖自然科学、工程技术、农业科学、医药卫生、经济管理、教育科学和图书情报等七大专辑。该数据库在高校一般有镜像站点或远程访问方式。

笔记

第二节 证据检索和收集的基本步骤

检索和收集证据的基本步骤如下(图 5-1):

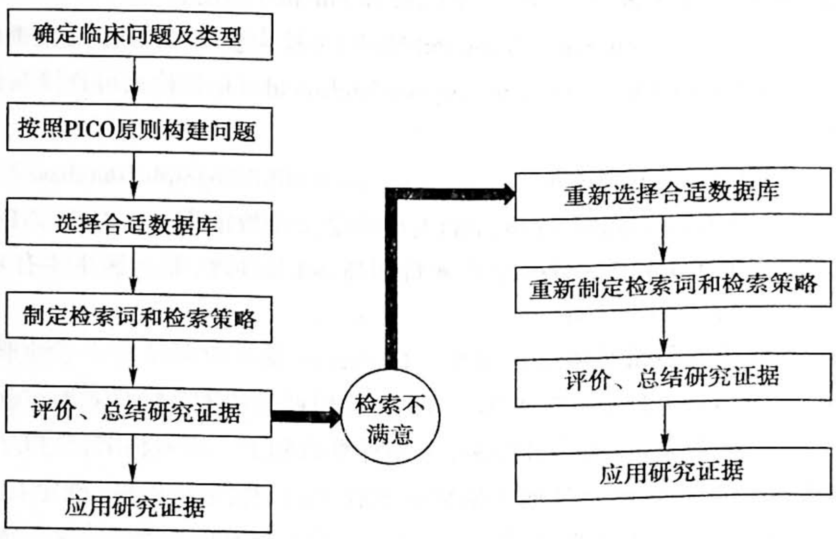


图 5-1 医学文献的检索策略

一、确定临床问题类型和构建临床问题

常见的临床问题主要涉及疾病的病因、诊断、预防、治疗、预后及不良反应,针对每一问题,根据需要了解的信息不同又分为背景问题("background" questions)和前景问题("foreground" questions)。背景问题是关于某种疾病、某一诊断技术或某一干预措施的一般知识,如心房纤颤是如何引起栓塞的?哪些人容易发生甲型流感?卡托普利是如何引起干咳的?前景问题涉及临床诊断、治疗的具体知识。循证临床实践过程中,提出的多数问题均为前景问题,如某诊断试验能否准确诊断某疾病?某干预措施的疗效和安全性如何?回答前景问题,需要明确问题涉及的 4 个要素:研究对象(patient)、干预措施(intervention)、与干预措施比较的措施(comparison intervention)和关心的结果(outcome),即根据 PICO 原则精心构建提出的前景问题。

明确临床问题类型及需要了解的信息、按照 PICO 原则构建临床问题,有助于正确选择数据库资源、合理制定检索词和检索策略。如背景问题我们可选择教材、专著和综述等,在 Cochrane 图书馆就不可能获得有关疾病病因、致病机理等方面的详细信息。而有关干预措施疗效,Cochrane 图书馆则是主要信息来源。

二、选择合适数据库

选择数据库,需要了解各数据库的特点、涉及的专业范畴和针对的临床问题类型。此外,尚需注意下列原则:

1. 选择计算机或网络数据库针对临床问题,可通过四种途径获得答案:咨询同事或专家、查阅教科书或专著、手工查阅相关杂志和计算机检索相关文献数据库。咨询同事或专家是快速、有效获取信息的方法,却受被咨询者知识更新程度的制约,常常存在一定偏倚;教科书或专著出版周期长,知识更新慢;手工查阅相关杂志能较快获得最新的信息,但非常费时;而计算机检索克服了时空障碍,只要掌握基本的技巧,既快速又高效。

2. 尽可能选择专业数据库综合性文献数据库如 MEDLINE、中国生物医学文献数据库(CBM)

等虽然覆盖了医学领域从基础到应用、各专业领域的资料，但有时却难以快速获得你真正需要的信息。因此，使用专业数据库更方便，易获得与专业直接相关的文献资料。但在缺乏专业数据库时，综合性文献数据仍然是最常用的信息资源。

3. 尽可能选择最佳文献数据库　最佳证据资源(best-evidence resources)是指采用明确的方法，对研究证据的科学性和临床相关性进行严格评价后建立的数据库。由于医学文献的层出不穷，临床医生应改变传统的文献查寻方法和技巧，从针对某一问题检索所有相关文献改变为按照"6S"模型检索文献，以便快速、高效获取最佳证据资源。根据"6S"模型，检索时应从证据系统(system)、证据集成(summaries)、系统评价摘要(synopses of syntheses)、系统评价(syntheses)、原始研究摘要(synopses of studies)和原始研究(studies)逐级检索，原则上如果从上一级数据库检索的文献解决了提出的临床问题，则无需继续检索下一级数据库。

三、选定检索词和制定检索策略

(一) 检索词

检索词的确定主要基于对提出的临床问题的分解单元，即 PICO 要素。通常检索词主要来源于 P(研究对象)和 I(干预措施)，而较少采用 C(对照措施)和 O(结果指标)。当根据 P 和 I 检索结果太多时，可考虑通过 C 和 O 进行限定。

(二) 检索策略

不同数据库，检索策略不全相同。检索策略就是采用逻辑运算符"AND"、"OR"和"NOT"将检索词进行组合的方式。

1. 扩大检索范围，提高查全率　当检索的记录太少时，可以使用以下方式提高查全率：

(1) 用主题词表(thesaurus)进行检索：如使用所选词的上位词进行检索、对主题词进行扩展检索、选用多个主题词检索、选用全部副主题词或对副主题词进行扩展检索、选用词表提示的相关词或以前的检索词进行检索。

(2) 用自由词检索：如果一个需检索的概念由几个自由词组成一个语句，应选用最能表达该概念的最少的自由词进行检索，因为一个语句中自由词的数量与检出的文献量成反比。

(3) 用"OR"运算符：用"OR"运算符可选择新的检索词，也可把同等或同义的检索词叠加组合起来进行检索，因此扩大了检索范围。

(4) 用截词：对检索词的词根或词尾加上截词符"*"进行扩展检索，可扩大检索词的范围，并防漏检。但用截词法耗费机时，而且容易产生假命中，应谨慎使用。

(5) 用通配符检索：用通配符"?"加在检索词中进行检索，可以检索出拼法不同而意义相同或相近的词，从而扩大检索范围。

2. 缩小检索范围，提高查准率　如果检索出的文献太多，可以用以下方法来缩小检索范围。

(1) 用主题词表进行检索：如选用主题词专指性不强，且该词下还有下位词，可选用下位词检索。

(2) 选准副主题词进行检索。

(3) 应用限定字段检索：限定字段检索时，常用的字段有 TI、AU、AD、PY、CP、AB、MESH、MJME、TG、NM、PT 等。如要提高通过主题词检索的准确性，可将检索结果限定在主要主题词字段进行检索。

(4) 用运算符：常用不缩小检索范围，提高查准率的运算符有 AND、WITH、NEAR、NOT 等。

完成检索策略后，针对选择的数据库进行检索。

四、判断检索结果

获得检索结果后，应判断所获信息能否回答提出的临床问题。如果不能获得满意答案，应

笔记

分析原因，是否数据库选择不当、是否检索词和检索策略制定不合理，还是确实该临床问题尚无相关研究证据。如果是从未经评价的数据库中检索的信息，尚需对检出的文献进行严格质量评价以确定其结果的真实性、临床重要性和适用性。

第三节 证据检索实例

一、证据检索实例一

（一）提出临床问题

糖尿病患者会发生诸多并发症，例如心血管疾病、肾脏疾病、眼底病变和神经系统疾病等。糖尿病患者发生心血管疾病和因心血管疾病死亡的风险明显高于非糖尿病患者。糖尿病患者合并高血压以后，容易发生眼底、肾脏、心脏、下肢血管和脑血管疾病。因此，糖尿病合并高血压患者的降血脂治疗很重要，血脂的干预水平要严于非糖尿病患者。但是否有研究证据支持这一观点呢？

临床病案 1：一名 55 岁男性患者，有 2 型糖尿病史 18 年，高血压病史 12 年，其血糖和血压水平一直控制良好。没有心肌梗死、心绞痛、脑血管意外和外周血管病史。最近复诊，血脂检查结果为：总胆固醇 5.2mmol/L，LDL- 胆固醇 3.0mmol/L，HDL- 胆固醇 1.6mmol/L，甘油三酯 1.8mmol/L。你似曾记得，某次讲座中讲授者说：对于 2 型糖尿病患者，即使血脂水平不高，降脂治疗也可预防心血管疾病的发生。考虑到降脂治疗的长期性及他汀类降脂药可能的不良反应。因此，你认为，在向患者推荐该治疗方案前，是否需要明确有无研究证据支持这种说法？

临床问题：血脂正常、无心脑血管病史的 2 型糖尿病患者，采用他汀类降脂药与安慰剂比较，能否预防心血管病的发生？

（二）构建临床问题

按照 PICO 原则分解上述临床问题（表 5-2），检出与上述临床问题直接相关的研究证据。

表 5-2 构建临床问题

P：患者及问题	血脂正常、无心脑血管病史的 2 型糖尿病患者
I：干预措施	采用他汀类降脂药
C：比较措施	安慰剂
O：结果	预防心血管病的发生

（三）检索相关研究证据

不同医学文献资料在设计、实施、统计分析、结果解释和论文报告等方面存在着差异，研究质量、结果真实性和可靠性及适用性也不同。因此，检索证据时，建议首先检索经他人评估和筛选过的循证医学资源，如果未检索出需要的信息，再进一步检索未经筛选的数据库。

1. 选择数据库

（1）首先检索经过评估或筛选的循证医学信息资源

- Clinical Evidence
- Best Evidence（Evidence-based Medicine and ACP Journal Club）
- Cochrane Library：Cochrane Database of Systematic Reviews（CDSR）
- UpToDate

笔记

- SUMSearch

（2）再考虑检索未经评估或筛选的综合信息资源

- PubMed
- Embase
- CBM

2. 确定检索词和检索策略

（1）检索词：根据构成临床问题的四要素，本病例检索可选择的检索词包括：type 2 diabetes、statin 和 cardiovascular disease。

（2）检索策略：上述临床问题采用检索词"type 2 diabetes"、"statin*"和"cardiovascular disease*"制定检索策略[type 2 diabetes] AND [statin*] AND [cardiovascular disease*]，并根据检索的数据库相应调整。

3. 检索相关数据库 我们检索了 Cochrane 图书馆，未发现 1 篇相关系统评价，只有 1 篇经济学评价的相关文献。再检索 PubMed，在"PubMed Services"下点击"Clinical queries"，输入检索词，在"Category"下点击"Therapy"，"Scope"下点击"narrow, specific search"，获得 38 篇文献（1966—2010 年 2 月 9 日）。最相关的为 Colhoun 等的文章："Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study（CARDS）: multicentre randomised placebo-controlled trial. Lancet 2004; 364: 685-96"。

4. 判断检索结果 获得检索结果后首先应该判断该结果能否回答之前提出的临床问题，对未经评价的文献，还需要进行严格的质量评价。

针对上述问题检出的 38 篇文献，仔细阅读题目和摘要，结合具体情况，发现 Colhoun 等的文章"Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study（CARDS）: multicentre randomised placebo-controlled trial. Lancet 2004; 364: 685-96"一文与提出的临床问题最相关。但此文献来自 PubMed，未经过质量评价，因此需要按照治疗性研究的质量评价原则分析此文献结果的真实性、临床重要性、可靠性和适用性（参见本教材第十一章），以确定能否正确回答上述临床问题。

二、证据检索实例二

临床病案 2：一名孕 25 周出生的早产儿，出生后 30 天胸骨左缘第 2 肋间闻及连续性杂音，脉搏增强。医生临床诊断为"动脉导管未闭（patent ductus arteriosus, PDA）"并已给予吲哚美辛治疗 PDA 一个疗程。婴儿非常虚弱且依赖呼吸机辅助呼吸，不便搬动做超声心动图检查以确诊患儿是否有 PDA。那么此时，能否根据体格检查结果（体征）能否诊断患儿的 PDA 呢？

（一）提出临床问题

动脉导管未闭是早产儿最常见的先天性心脏病，若未及时诊断与处理，常可诱发或促进充血性心衰、慢性肺疾病、颅内出血和坏死性小肠结肠炎（NEC）等的发生发展。足月儿 50% 在生后 24 小时内，90% 在 48 小时内，几乎 100% 在 72 小时内动脉导管（DA）关闭，在生后 1~3 个月常达到完全解剖上关闭。但若胎龄≤29 周，PDA 发生率在 0~24 小时为 80%，24~48 小时为 40%，48~72 小时为 7%。早产儿一旦确诊为持续性 PDA，建议尽早治疗。药物（吲哚美辛等）和手术（PDA 结扎）治疗 PDA 安全、有效。因此，应尽早确诊，减少并发症发生。

超声心动图诊断 PDA 最敏感准确，临床体征较迟发生，但对判断与 PDA 有关的远期疾病发生关系更密切。常见临床体征包括连续性杂音、心前区搏动增强、水冲脉、脉压增大或存在进

笔记

行机械通气的指征。体征不同诊断意义也不同。作为标准诊断方法的超声心动图虽然准确，但该患儿病情危重、不便搬动，而床旁超声心动图机并不是任何医院均有。如果临床体征能作为筛查手段，将有助于早期诊断和治疗。尽管临床体征广泛用于 PDA 的初步诊断，其准确度到底如何呢？能否有助于诊断上述早产儿的 PDA 呢？

临床问题：对依赖呼吸机辅助呼吸的极低体重（体重 < 1000g）早产儿，临床体征诊断 PDA 的准确性如何？

（二）构建临床问题

按照 PICO 原则分解上述临床问题（表 5-3），检出与上述临床问题直接相关的研究证据。

表 5-3 构建临床问题

P：患者及问题	极低体重早产儿
I：干预措施	临床体征
C：比较措施	超声心动图（金标准）
O：结果	诊断动脉导管未闭

（三）检索相关研究证据

不同医学文献资料在设计、实施、统计分析、结果解释和论文报告等方面存在着差异，研究质量、结果真实性和可靠性及适用性也不同。因此，检索证据时，建议首先检索经他人评估和筛选过的循证医学资源，如果未检索出需要的信息，再进一步检索未经筛选的数据库。

1. 选择数据库 目前尚无专门针对诊断试验证据的数据库，只能通过综合性数据库检索诊断试验证据。

（1）首先检索经过评估或筛选的循证医学信息资源

- Best Evidence（Evidence-based Medicine and ACP Journal Club）
- Cochrane Library：Cochrane Database of Systematic Reviews（CDSR）
- UpToDate
- SUMSearch

（2）再考虑检索未经评估或筛选的综合信息资源

- PubMed
- Embase
- CBM

2. 确定检索词和检索策略

（1）检索词：根据构成临床问题的四要素，本病例检索可选择的检索词包括：patent arterial duct、diagnostic test、clinical examination、sensitivity、specificity、preterm。

（2）检索策略：上述临床问题采用检索词 patent arterial duct、sensitivity、preterm 制定检索策略[patent arterial duct] AND [sensitivity] AND preterm，并根据检索的数据库相应调整。

3. 检索相关数据库 首先检索二次文献数据库 Best Evidence 和 CDSR，未检出相关文献。

再检索 PubMed（http://www.ncbi.nlm.nih.gov/PubMed/），从页面左侧的"Clinical Queries"进入检索口"Search by Clinical Study Category"，输入"[patent arterial duct] AND [sensitivity] AND preterm"，在"category"下选择"diagnosis"，在"scope"下选择"broad，sensitive search"，检出 33 篇相关文献。

4. 判断检索结果 获得检索结果后首先应该判断该结果能否回答之前提出的临床问题，对未经评价的文献，还需要进行严格的质量评价。

针对上述问题检出的 33 篇文献，仔细阅读题目和摘要，结合具体情况，发现"Davis P，Turner-Gomes S，Cunningham K，et al. Precision and accuracy of clinical and radiological signs in premature infants

笔记

at risk of patent ductus arteriosus. Arch Pediatr Adolesc 1995; 149(10): 1136-41"一文与提出的临床问题最相关。但此文献来自 PubMed，未经过质量评价，因此需要按照诊断试验的质量评价原则分析此文献结果的真实性、临床重要性、可靠性和适用性（参见本教材第九章），以确定能否正确回答上述临床问题。

（李　静　吕　明）

第六章　医学研究文献的评阅

学会阅读和评价医学研究文献是对临床医生和医学生职业素养的基本要求。这是因为，临床医生所从事的医、教、研工作，与临床研究息息相关，需要对医学研究文献进行阅读与评价。如在临床实践中，为避免知识老化和保持临床技能水平不降低，特别是当遇到自己难以解决的诊断、病因、防治、预后问题时，除了咨询同事或向上级医生请教外，更多的则是通过检索、阅读与评价医学研究文献，从中寻找答案；在临床教学过程中，同样需要及时掌握本学科的新知识、新进展，以更好地向学生传授专业知识；在临床科研的选题立题、提出科学假设过程中，也需先期进行大量的文献复习，以充分掌握相关研究领域的历史、现状及存在问题，由此，提出立题依据、明确研究重点。然而，临床医生的工作性质决定了他们不可能有过多的时间去阅读相关研究文献，这就要讲究一定的文献阅读方法和技巧，通过提高阅读效率，完成必要的文献阅读以及知识的更新与储备。因此，掌握正确的医学文献阅读与评价方法，科学有效地管理好医学专业文献，将有助于临床实践、教学和科研，同时也是临床医生实现自我终身学习、保持与提高临床技能水平的必由之路。

第一节　阅读与评价医学研究文献的重要性

一、医学研究文献的特殊性

（一）数量庞大

医学研究文献比较特殊，首先医学文献数量非常庞大且仍处于快速增长之中，已位居各学科文献量之首。同时医学文献还具有载体多、发表语种多的特点，以及重复发表、发表分散以及滞后发表等一系列问题。例如，全球范围内生物医学期刊已有数万种之多，每年发表文献达几百万篇以上。此外，还有大量的电子文献出版物、未公开发表的文献以及灰色文献等，从而形成了海量的医学研究文献以及医学信息资源。学会正确地阅读和评价医学文献，其重要性不言而喻，因此，医学生和临床医生应具备一定的检索、阅读以及评价医学研究文献的能力，学会从中汲取有用的知识和信息。

（二）医学文献的分类体系复杂、质量参差不齐

文献是指以文字、图像、公式、视频与音频、代码等形式，将信息、知识记录或描述加以存储、传播的一切载体。医学文献就是记录有医学相关知识或信息等载体的总称。

1. 按照载体的属性可分为书写型、印刷型、电子型、微缩型和视听型等文献类型。其中书写型文献专指手工书写与抄写的文献，如病历、实验原始记录、设计草图等。印刷型文献特指纸制出版物，为图书馆收藏的主要类型，种类繁多，包括医药图书（专著、教科书、工具书）、生物医学期刊、学位论文、会议文献、研究报告或官方出版物（如卫生年鉴）、专利文献、医疗器械技术标准以及产品资料、医疗技术档案等。电子型文献是指医学信息以光盘、网络、软盘等形式存储和传播，目前已成为主要文献主体，且有逐步取代印刷型文献之势。

2. 按照研究内容和组织形式，医学文献又可分为一次文献、二次文献、三次文献以及零次文献等。其中一次文献又称原始研究文献，是基于作者本人的经验总结或者科研成果而创作的、具有一定原创性的一类文献。如生物医学期刊中最为常见的论著（article），是典型的一次文献，数量庞大，已达到数千万篇之多；为方便检索、利用这些原始研究文献，国内外一些医学信

笔记

息研究机构或组织，基于一次文献的外部特征进行收集、整理、压缩、归类，并按照一定顺序组织编排等初步加工过程，生产出一系列的二次文献，如书目、索引、文摘、题录等。这些二次文献具有汇集性、工具性、综合性、系统性等特点，出版形式包括印刷型和电子型等，其中医学文献检索数据库，以 MEDLINE、Embase、中国生物医学文献数据库（CBM）、CNKI、VIP 等，最为常用。这些生物医学文献数据库，均由专门机构负责收录、整理、加工全球范围内发表的各类原始研究文献，收录条目数量庞大，但彼此间互有交叉、质量也参差不齐；三次文献是在充分利用二次文献的基础上对一次文献做出系统整理和概括，进而汇总编写而成的综述性文献。主要包括综述、年鉴、手册、百科全书、文献指南等。其中以系统评价（systematic review）、临床实践指南（clinical practice guideline，CPG）等最为常见；零次文献是那些未经正式发表或未进入社会交流的最原始文献。如设计草图、实验记录、草稿、会议记录、内部档案等。

3. 按照证据质量分类　近年来，随着循证医学的引入和推广，如何获取证据，特别是高质量的证据，成为临床医生实践循证医学的关键，而有重要临床价值的高质量医学文献则是证据的主要来源。为此，加拿大临床流行病学与医学信息家 R. Brian Haynes 提出了 6S 证据模型。该模型是由最早的 4S 模型、5S 模型衍化而来。6S 类证据依次为系统类证据（systems，如计算机决策支持系统 CDSS）、指南类证据（summaries，如循证临床实践指南 CPG）、集成类证据（synopses of syntheses，如 DARE、提要类循证医学杂志系列等）、系统综述类证据（syntheses，如 Cochrane 系统评价 / 综述）、原始研究提要类证据（synopses of studies，如 handbook of clinical practice），原始研究级证据（studies，如 ACP JC+ 等）。这 6 类证据均由专业组织或机构分别从专业和方法学质量角度进行严格评价后经筛选而来的证据，证据质量一般较高。因此，按照证据质量和级别可将医学文献分为 6S 类证据文献和非 6S 类证据文献。需要注意的是，6S 类证据因入选标准较高、筛选制作要求较为苛刻，同时耗时耗力，生产数量比较少，临床实践中有关病因、诊断、治疗、预后问题，多数情况下并无现成的 6S 类证据可供参考，仍以传统的、未经严格评价的、非 6S 类医学文献为主。

证据级别的分布一般呈现金字塔状，高级别证据是由低级别文献信息综合提炼加工而成，数量少，分布在塔尖；低质量文献或证据数量庞大，分布在塔底（图 6-1 左图）。例如，循证临床实践指南是指导疾病防治的高级别证据，综合萃取了系统评价及 RCT 等证据信息，在数量上要少于 RCT 等原始研究文献。

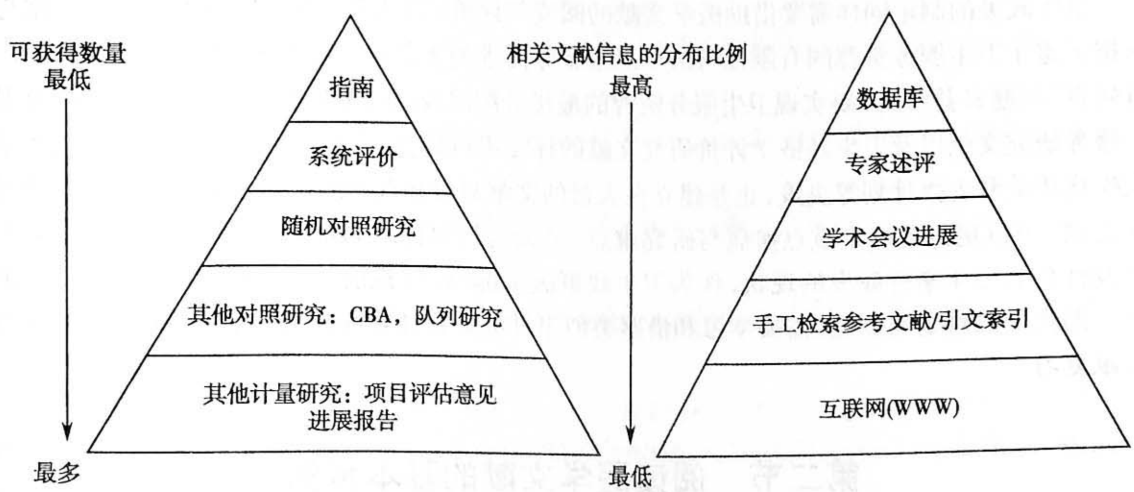


图 6-1　医学文献与信息资源分布金字塔图

二、阅读和评价医学文献的重要性

（一）医学文献与临床医疗实践

在日常的临床实践中，要求临床医生掌握临床诊治新技术与新方法，离不开对医学文献的

阅读与评价。特别是现在倡导循证医学实践，要求高素质的临床医生结合医疗环境和患者的主观意愿，将最佳证据应用到具体的临床实践之中，从而创造一流的医疗水平，更好地为提高国民健康水平服务。最佳证据来源于高质量的诊断、防治、预后以及病因学 / 危险因素研究文献。另外，随着病人维权意识的提高，病人及家属也可能利用一些现成的检索资源和检索工具，查阅医学文献，希望掌握疾病诊治相关的知识与信息，并从中找出一些“证据”主动与医生探讨。鉴于医患关系紧张已是不争的事实，这也迫使临床医生及时掌握相关文献信息，以利于医患间的交流与沟通，创造和谐的医患关系。

（二）医学文献与临床科研

医学科研，选题立题是基础。当今社会是“知识爆炸”的时代，由于科技与信息科学的高度发展，信息量本已十分庞大，而新的海量信息又源源不断的涌现，这在医学相关领域表现得尤为突出。在此背景下，临床医生担负着临床医疗与临床科研的双重使命，临床研究要有所创新，需要进行大量的文献复习，以掌握最新的专业进展与研究动态，这也离不开对医学文献的评阅。

（三）医学文献与医学教育

在医学教育的不同阶段，对医学文献有着不同层次的需求。例如，处于培训阶段的临床医学生，对医学文献阅读、评价的需求不高，很多入门问题已由专人帮助解决；但过了该阶段，在独立进行临床实践活动中，所遇到一系列临床问题，这就要通过阅读和评价医学研究文献，大多由自己加以解决。

目前，由于医疗新技术与新方法不断涌现，知识更新周期明显缩短，过去认为是最佳的技术方法，逐渐被淘汰。临床医生要保持和提高临床技能水平，需要进行终身教育，不断学习与更新专业知识。其中阅读专业文献，成为大多临床医生知识更新的首选。例如有研究表明，要维持业务水平不落伍，需要定期阅读大量专业文献，以及时掌握本学科新进展，阅读的杂志越多，则越有可能追踪到全部相关文献。如果阅读 20 本专业杂志，基本可以覆盖 80% 的相关文献，若要实现全覆盖，则至少需要订阅 60 本专业杂志。这对临床医生是一个严峻挑战，为此，只有学会正确的阅读与评价医学文献，方能实现自我终身教育，使自己的知识水平永葆一流。

（四）医学文献与卫生决策

卫生政策的制定同样需要借助医学文献的阅读与评价，以从中发现决策所需的重要证据与依据。鉴于卫生服务资源的有限性与医疗卫生服务需求的无限性的矛盾将长期存在，“看病难、看病贵”问题日益突出，要实现卫生服务资源的最优分配以及医疗卫生服务的效率最大化，对卫生服务研究文献以及卫生经济学评价研究文献的评阅不可缺少。同时，在国家层面上形成的重大疾病攻关和支撑计划等决策，也是建立在大量的文献复习和调研的基础之上的。通过阅读医学文献，可以更好地确定重点疾病与研究重点，从而有针对性地制定招投标指南。此外，对于突发性公共卫生事件频发的现状，作为卫生政策决策部门，要提前制定处置预案，以防患于未然。而预案的规划与制定，需要学习和借鉴类似事件的处置经验和教训，同样也离不开大量的文献复习。

第二节　阅读医学文献的基本步骤

一、医学文献检索

文献检索分为手工检索和计算机检索。随着计算机技术和通信技术的迅猛发展，使得计算机文献检索得以广泛应用，由于其检索速度快、范围广、内容新、检索入口多、功能强且用户使用方便，已成为医学生必须掌握的关键技能之一。

笔记

（一）医学文献来源与检索资源

医学文献具有来源广、种类多、分布广泛等特点，且有一定规律可循。例如，按照文献可及的数量以及分布密度顺序排列，一般呈现金字塔状分布（图 6-1 右图）。其中电子文献检索数据库（电子文献索引或全文数据库），收录了包括指南、系统评价、临床原始研究以及其他计量研究等绝大多数医学研究文献。目前常见的中英文生物医学文献数据库中，主要包括 MEDLINE（美国“医学索引在线，或 PubMed”）、Embase、中国生物医学文献数据库（Chinese Biomedical Literature Database，CBM）、CNKI、VIP 等。当然一篇医学文献有可能被多个电子文献数据库同时收录，在检索文献时，可能被多次检出，需要进行剔重。

尽管电子文献数据库是医学文献的主要来源，但仍存在一些不足。如大多只能收录摘要，未能收录文献全文。其次可能存在发表性偏倚，阴性结果文献数量可能被严重低估。另外文献的收录也存在地域性发表偏倚。如在一个收录了 3000~4000 种杂志的文献数据库中，只有 2% 的文献来自中低收入的发展中国家，而发表在这些国家国内杂志上的文献超过 90% 并没有被主要电子检索数据库收录。因此，为弥补电子文献数据库的上述缺陷，常利用互联网（WWW）查阅一些灰色文献作为常规医学文献检索的重要补充。灰色文献是指由官方、NGO、学术机构等非赢利单位出版发行的印刷品和电子出版物。形式多种多样，既可以是技术报告、研究报告，也可以是技术文件、技术说明与标准，以及一些非赢利性的译文、著作、官方文件等。因这些文献信息多未经过严格评价，称之为灰色文献。除此之外，互联网上的灰色文献还包括数量庞大的、有潜在价值的网页信息，多由一些小的或不知名的机构、非政府组织（NGOs）、学术研究中心、咨询公司等创建。但从这些海量的灰色文献资源中提取有价值的信息作为文献检索的重要补充，还需借助于一些数据挖掘与知识发现技术。

（二）文献检索基本步骤

无论是人工检索还是计算机检索，都大致分为以下几大步骤：从分析检索题目入手，明确检索要求；进而选择检索工具，制定检索策略。如一般将那些文献类型全、数量大、时差短、途径多、著录标准的检索工具作为首选；最后选择检索途径，检索并获取原始文献。

1. 检索问题的提出 围绕研究目的，将特定的研究问题进行规范化和标准化，形成结构化检索问题，进而按照 PICOS 原则解析后，制定检索策略，进而选择检索数据库，系统检索文献。根据 PICOS 原则，检索问题一般可解析为以下 5 大要素：

“P”：为 patient（病人）或 population（群体）的缩写，表示他（她）或他们患的是什么病、存在什么需要解决的临床问题。

“I”：为 intervention（干预措施）的缩写，表示针对病人存在的临床问题，拟探求的干预措施是什么？

“C”：为 comparison（比较）的缩写，表示设置的对照比较措施是什么？如安慰剂或其阳性对照等。

“O”：为 outcome（结果）的缩写，表示与干预措施相关的最终结局是什么？如不良事件发生率、病死率等。

“S”：为 study（研究类型）的缩写，表示具体的研究设计方案是什么？如原始研究类型主要包括随机对照临床试验、交叉试验、非随机同期对照试验、队列研究、前 - 后对照研究、病例对照研究、病例系列或报告等。

由于检索问题各具特点，检索者不一定将关注的问题均解析为上述 5 项，也许只需 2~3 项即可，但上述 5 要素的解析思路，有利于检索者形成一个完整的检索路径。

2. 制定检索策略 检索策略（search strategy）是指在解析相关问题的基础上，结合检索目的和信息需求，选择检索系统，确定检索词，构造检索式，从而制订出较为完善的检索计划或方案。

笔记

其中构造检索式需要使用检索系统规定或允许的符号(运算符),并用于连接已确定的检索词。检索系统中的运算符有位置运算符、逻辑运算符、限制运算符、截词符等。常使用的逻辑运算符有:①"AND"(逻辑"与"):其作用为缩小检索范围,提高查准率;②"OR"(逻辑"或"):其作用为扩大检索范围,提高查全率;③"NOT"(逻辑"非"):其作用为缩小检索范围,提高查准率。

3. 选择检索数据库、系统检索文献 选择检索数据库,确定检索方法。一般将那些文献类型全、数量大、时差短、途径多、著录标准的检索数据库作为首选;同时为提高检索效率,在检索内容与顺序安排上,有一定的讲究。一般是先寻找可靠的三次文献,如指南、系统评价等,这些文献综合了大量相关的原始研究结果,且经过了加工和提炼。若无这样现成的文献,再寻找可靠的原始研究文献。

开始使用电子文献数据库时,最好寻求图书管理员的帮助,以尽快熟悉检索方法,提高检索效率。

二、阅读医学文献的基本步骤

医学研究文献具有两种属性,一是外部特征,比如题目、作者、作者单位、发表期刊、卷、期、页码、年限、语种等,这些信息由于具有唯一性和指向性的特点,常被用来标识文献,供文献的收集、整理、存储、传播与查询之用;二是内部特征,包括研究目的、科学假设、设计、研究方法、研究对象、主要结果、讨论和结论等方面内容,既是文献的核心,又是文献阅读的重点(图6-2)。

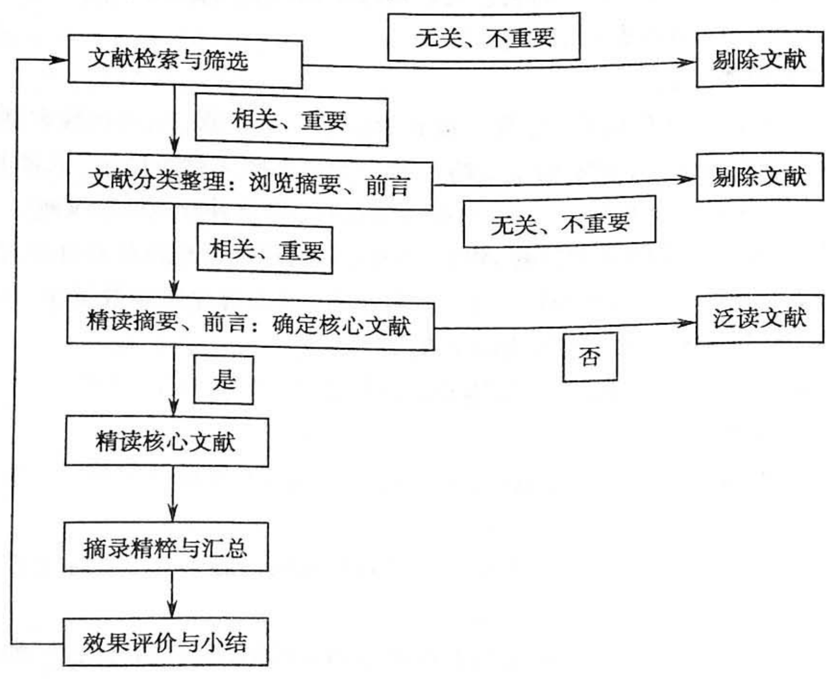


图6-2 文献阅读的基本流程图

(一)结合阅读文献的目的,筛选、甄别、获取相关医学研究文献

开始阅读之前,一定要弄清楚"为什么要阅读文献?""要阅读哪些文献?"带着问题有针对性地筛选、甄别、获取医学研究文献。筛选文献不能无的放矢,根据具体问题,一般先从阅读文献摘要入手,按照与阅读目的的关联性大小以及文献的时效性,依次安排文献阅读的先后顺序,从中遴选出最新、关联程度最高的文献,进而阅读全文。这一点对提高阅读效率尤为重要。鉴于候选文献中的摘要,无论是结构式摘要抑或非结构式摘要,提供的信息量有限,有时难以取舍,需要进一步获取全文,以确认是否最终纳入。

笔记

（二）熟悉文献的基本结构、选择阅读方式和阅读重点

对于纳入的文献，应首先熟悉其基本结构，确定阅读重点。因个人精力和时间有限，最好在正式阅读开始之前，对检索到的文献进行分类和整理，根据个人精力、时间以及与个人关注问题的密切程度，将纳入文献分为：精读文献和泛读文献。如，对于与个人关注问题密切相关的文献可采用选择性精读文献的方法；而对于与个人关注问题关系不密切的文献可以采用泛读的方法进行。在医学研究领域中，重要文献是指那些对学科发展具有里程碑意义的原始研究文献以及学术权威撰写的综述或述评等。对此类文献应设法获取全文并仔细阅读。其他文献以泛读为主，先阅读题目、摘要，在此过程中，若发现文献有价值可升格为重要文献，再精读全文。若同类文献较多时，考虑精读文献的顺序，新近且重要的文献，优先进行精读。

一篇完整的原始研究文献一般包括摘要、前言、材料和方法（或对象和方法）、结果、讨论（包括结论）和参考文献六部分。阅读目的不同，对同一篇文献的阅读重点有所侧重。如果想了解该文献的结论是否适合于自己的病人，可直接阅读“材料和方法”部分了解其设计方案、病例的选择标准等以判断其结论的应用范围，无需从头读到尾。若为了解学科新进展，应重点阅读文献的“结果和结论”部分；若为启迪研究思路，则应以文献的研究方法为阅读重点，可取长补短、突破固有思维模式，发现创新点；若为临床科研的选题及立题提出依据，先阅读文献的“前言”部分，掌握该研究领域的历史与现状，同时在该文献“讨论”部分查找不足之处、方法学缺陷等，这些就是将来的研究方向与创新点。

（三）阅读文献、摘录文献精粹

对纳入的文献，即使一篇与自己研究目的高度相关的文献，也不可能全盘照搬，需要从中摘录出精粹部分，可以是文献中的某段话、某种方法，也可能是一幅图表等，宜精不宜多，过多内容会显得杂乱无章，对初学者无益。在摘录文献精粹的基础上，进一步加以系统总结。逐一将文献的精粹部分，加以汇总；为方便管理，需要创立一个文档文件，将相关内容作长久保存。同时，在汇总过程中，最好能将文献的读后感，包括阅读后有哪些收获、新启发、新思路等提炼加工后，一并纳入。

一般的文献泛读可以在5~15分钟内完成，精读文献则需要2小时到1周时间不等。相对文献泛读，精读文献时需要摘录的文献信息更为全面细致。特别是应弄清以下基本问题：

1. 研究目的是什么 通过阅读前言部分，明确该文献的科学假设、拟解决的关键问题、立题背景和依据是否充分等。

2. 研究方法与研究对象是什么 在这部分，需要了解具体的研究设计方案、研究方法是否新颖合理、样本来源与样本大小如何、设置哪些测量指标以及指标的实际价值和意义如何？

3. 主要结果和重要发现有哪些 全面熟悉文献的主要结果，明确文中的新发现、新贡献，用于结果报告的重要图表有哪些？

4. 讨论与结论 文献的结论是什么？该文的主要结果是否支持这些结论？文中的立题、设计、方法和讨论间是否有内在的逻辑关系？研究目的是否实现？主要不足有哪些？还遗留有哪些问题未解决？

5. 小结 精读后，最好能加以小结。如与同类文献相比，该文献有什么共同点和不同点？作者的整体思路有无创新之处？必要时，可进行类似SWOT分析，找出优势和不足。

第三节　医学研究文献的评价要素与方法

一、确定医学研究文献评价的范畴和内容

评价文献时同样需要带着问题，有针对性地进行。评价目的不同，决定了评价范畴和重点。

笔记

文献质量评价涉及两个方面的内容，即报告质量和方法学质量。其中报告质量是指文献报告内容的全面性和完整性以及和相应报告规范的符合程度。方法学质量是指文献制作过程中遵循科学标准、有效控制混杂与偏倚、使研究结果达到真实可靠的程度，为文献质量评价的核心内容。

鉴于评价往往具有很强的主观性，受评价者自身能力与水平的限制，同一篇文献，其评价结果有可能存在较大出入。因此，文献评价方法的选择以及评价过程是否规范透明，尤为重要。临床流行病学作为临床医学的一门基础学科，强调应用科学的方法学强化临床科研设计、排除各种偏倚、混杂因素的影响，确保研究结果的真实性和研究结论的可靠性，使得科学研究获得的成果能够用于指导临床实践、教学与科研。因此，临床流行病学的相关质量标准已成为质量评价的参考依据，并在国际上获得公认。因此，为保证评价质量，评价者应学习和掌握临床流行病学中有关研究设计、测量与评价的基本原则和方法，有助于培养批判性思维能力。

评价内容主要包括真实性评价、重要性评价及适用性评价，用以依次回答下列问题：该研究结果本身是否真实可靠？有多大临床意义和实用价值？用于临床实践的可行性及适用程度如何？其中研究结果的真实性和重要性是评价的重点，只有真实、可靠、重要的研究结果才有利用价值。

二、评价医学研究文献的一般原则

阅读医学研究文献的目的，主要是回答“what”类问题，以全面了解该文献中的研究目的、对象、研究方法、主要结果和重要结论等内容；而评价医学研究文献，则是回答“why”类问题，一般选在二轮精读时进行，要求评价者具备批判性思维并掌握一定的评价原则与方法。

（一）真实性评价原则

评价真实性的关键，在于考核研究过程中是否有效控制了混杂与偏倚对结果的影响。这些因素对结果的影响大小决定了真实性的程度。因此，评价真实性应综合考虑研究结果是来自何种设计方案、有无对照组以及设置是否恰当、研究对象的诊断标准是否可靠、纳入 / 排除标准如何、样本量是否足够、组间重要的基线状况是否可比、有无相关偏倚因素存在以及是否采取了相应的防止或处理的措施、依从性如何、对相应的试验观测指标及资料所采用的整理、统计分析方法是否恰当等。如针对定量研究文献，JAMA 用户指导手册中专门制定了 5 项指导性评价原则：①研究对象是否与研究问题有关？②研究对象的选择是否有充足的理由？③数据资料的收集方法是否与研究目的与场所匹配？④数据资料的收集是否完整充分、足以描述观察事件？⑤资料的分析是否合适以及发现的结果是否被充分证实？等等。这 5 项原则对一般定量研究文献的评价均有一定的参考价值。

（二）重要性评价原则

评价临床研究结果的重要性常借助于一些定性或定量指标。如在临床试验中，定性指标有事件发生率（如病死率、生存率、治愈率……）、绝对危险降低率（absolute risk reduction，*ARR*）、绝对获益增加率（absolute benefit increase，ABI）、相对危险度降低率（relative risk reduction，*RRR*）、相对获益增加率（relative benefit increase，*RBI*），需治疗多少例患者才能获得一例最佳效果（number needed to treat，*NNT*）以及需治疗多少例患者才能发现一例不良反应（number needed to harm，*NNH*）等。对于定量指标，则较为单一，主要计算组间均数差值。重要性包括临床重要性与统计学意义两个方面，两者应相互结合，作综合评价。统计学意义的判定可以通过假设检验和区间估计加以实现。若假设检验的 P 值小于预先设置的检验水准（常设为 0.05），则可认为组间差异有统计学意义。这里要强调的是组间差值的大小与 P 值无必然联系，假如组间差值无临床意义，P 值再小，也无临床应用价值。当某种研究结果既有临床意义，又有统计学意义时，即能作出肯定性的结论；如仅有临床意义而无统计学意义时，不能盲目否定其临床价值，应计算Ⅱ型错误率或检验效能加以核实；若文献结果既无临床意义，又无统计学意义，则此类文献的重要性可忽

笔记

略。临床重要性的判断，还应做卫生经济学的评价，进行成本-效果（cost-effectiveness）、成本-效益（cost-benefit）以及成本-效用（cost-utility）分析，使那些成本低、效果佳的研究成果，得以推广应用。

（三）适用性评价

适用性评价同样需要结合阅读文献的目的，是为了指导临床实践、教学，还是为临床科研所用。若将上述真实性好且有重要临床价值的文献结果在临床实践中加以应用和推广，应结合自己病人的实际病况和接受程度、现有医疗条件和知识技能水平，以及社会经济状况的承受能力等，对其临床适用性展开评价。鉴于当前高质量的临床研究文献多来源于发达国家，而由于人种、社会环境、经济水平、医疗条件乃至生物因素在国家层面上差异较大。因此，评价适用性，更要结合不同的国情、种族以及病人特点，切不可生搬硬套。要对具体的问题作具体的分析，方可作出是否适用的决策。

若在临床科研、教学过程中，为掌握学科最新进展或发展方向、拓展新思路、新视野之用，也要进行适用性评价，对文献结果是否具有外部真实性，即能否将结果推广应用到研究对象以外的群体或环境时，应考核研究人群与其他人群的特征差异、研究对象类型以及社会环境和经济等因素是否会影响适用性等。

三、临床研究类型及其评价工具的合理选择

为方便评价，针对不同研究类型，已有一些现成的严格评价工具或标准，可供参考借鉴。评价工具大多由一些知名学术机构或组织研发。例如 JAMA 发布的用户指导手册系列、CASP 严格评价技巧项目网提供的系列质量评价标准等，专门用于评估包括系统评价、随机对照试验、病例对照研究、队列研究、描述性研究、诊断试验和经济学评价研究等在内的不同类型文献。评价工具可分为清单类（checklist）和尺度评分类（scale）两种。

不同的研究设计，其科学论证强度不同，评价方法及评价工具也有所不同。

1. 原始研究评价工具 以随机对照试验的评价工具最为多见。由于随机对照试验采用了随机、盲法以及设置对照组，最大限度地控制了混杂和偏倚对结果的影响，确保了结果的真实性，被认为是一种论证强度较高的设计方案，因而在临床研究中备受推崇，相关文献发表也很多，成为临床证据的重要来源，相应的质量评价方法也发展很快。报告质量的评价可借助于 CONSORT（Consolidated Standards of Reporting Trials），方法学质量评价可选择的工具比较多，如 Cochrane 手册中的偏倚风险评价工具，Jadad 评分等；对于观察性研究（包括队列研究、病例对照研究等），报告质量评价可以选择 STROBE（Strengthening the Reporting of Observational Studies in Epidemiology，STROBE），方法学质量评价工具也比较多，常见有 CASP 工具（Critical Appraisal Skills Programme，CASP）和 NOS 评分（The Newcastle-Ottawa Scale，NOS）等。

2. 二次研究评价工具 对二次研究文献可评价其方法学质量以及报告质量。如评价系统评价的方法学质量工具有 OQAQ、AMSTAR 等；有关报告质量评价工具有 QUOROM 及其升级版（PRISMA）等。其中，PRISMA 是用来规范系统评价的一种标准报告格式，包括 27 个条目，覆盖了系统评价中的摘要、简介、方法与结果等方面内容，可用来与一篇系统评价内容逐一进行比对，考核其报告内容是否完整。

（康德英）

笔记

第七章　医学研究证据的系统评价与meta分析

随着医学科学日新月异的发展、研究新领域的日趋开拓、人们认识的不断深化，临床医学研究层出不穷，为决策者提供了大量科学信息。医务人员和研究者为了获得新知识、新观点和新技术以扩展视野，提高学术水平，需要阅读大量文献。但现有临床研究多数规模较小，纳入研究对象数量有限，针对同一种疾病的同一或同类干预措施的文献资料有时数量较多，质量良莠不齐，结论也不尽一致。如何从浩如烟海的医学文献信息中快速、高效率地获得所需资料，以进行科学决策，已成为我们面临的巨大挑战。这就需要有一种方法、一种标准将全面、系统收集起来的文献，去粗取精、去伪存真、综合分析，从而得出一个真实性及可信性较高的结论，以供读者参考应用，这就称为文献综述；如针对某一具体临床问题，采用一套规范、科学的方法全面收集、认真选择、严格评价和科学分析相关研究资料，得出综合可靠的结论，此乃系统评价(systematic review)。

临床医生工作繁忙，要想为解决一个临床问题，去查遍相关文献并进行综合评价，难度很大，可是，假如您面对一个实际问题需要您科学回答，这就是挑战：如一个妇产科医生，面对有早产危险的孕妇，而早产婴儿往往有可能胎内发育不全，特别是因肺发育未成熟，死亡率和呼吸窘迫综合征的发生率都非常高。有专家提出，如果对可能早产的孕妇使用激素，可促进胎儿的肺发育。为了明确对可能早产孕妇使用激素后是否能减少早产儿的死亡率和呼吸窘迫综合征的发生率，你查寻了有关资料，发现有7个高质量的随机对照试验，其中5个试验结果为阴性(使用激素后未能减少早产儿的死亡率和呼吸窘迫综合征的发生率)，2个试验结果为阳性，你将作何决策呢？科学回答这个问题就需要对这7篇文献进行系统评价，以助自己的科学决策。

因此，本章将重点阐述系统评价及meta分析的方法及其相关的质量分析和评价原则，以供读者回答这一挑战的参考和应用。

第一节　系统评价概述

一、基本概念

(一) 系统评价

系统评价是一种全新的文献综合方法，指针对某一具体临床问题(如疾病的病因、诊断、治疗、预后)，系统、全面地收集现有已发表或未发表的临床研究，采用临床流行病学严格评价文献的原则和方法，筛选出符合质量标准的文献，进行定性或定量合成(meta-analysis，meta分析或荟萃分析)，得出可靠的综合结论。系统评价可以是定性的(定性系统评价，qualitative systematic review)，也可以是定量的(定量系统评价，quantitative systematic review)即包含meta分析过程，系统评价的整个过程非常明确，使其具有独特的优点——良好的重复性。

系统评价可为某一领域和专业提供大量的新信息和新知识，多数是可信的。但是，由于是对原始文献的二次综合分析和评价，受原始研究文献的质量、系统评价的方法及评价者本人的专业知识、认识水平和观点等的制约，因此，读者在阅读系统评价的观点和结论时，一定要持谨慎的态度，不能盲目被动地接受。

(二) Cochrane系统评价

笔记

Cochrane系统评价是Cochrane协作网的评价人员按照统一工作手册(Cochrane reviewers'

handbook），在相应 Cochrane 评价小组编辑部的指导和帮助下所完成的系统评价。由于 Cochrane 协作网有严密的组织管理和质量控制系统，严格遵循 Cochrane 系统评价者手册，采用固定的格式和内容要求，统一的系统评价软件（Review Manager，RevMan）录入和分析数据、撰写系统评价计划书和报告，发表后根据新的研究定期更新，有着完善的反馈和完善机制，因此 Cochrane 系统评价的质量通常比非 Cochrane 系统评价质量更高，被认为是单一的、评价干预措施疗效的最好证据资源（best single source）。

目前，Cochrane 系统评价主要针对研究疾病防治、康复疗效和安全性的随机对照试验进行评价，其方法较完善和规范，诊断性试验的系统评价已开始进行。

（三）meta 分析（meta-analysis）

meta 分析由心理学家 Glass 1976 年首次命名，国内翻译为荟萃分析、汇总分析。就其定义目前仍然存在不同的争议。Huque 及多数专家认为："meta 分析是一种统计分析方法，它将多个独立的、可以合成的临床研究综合起来进行定量分析。"因此，如果没有明确、科学的方法去收集、选择、评价临床研究资料，而仅单纯采用统计方法将多个临床研究进行合成并不能保证结论的真实性和可靠性。

目前系统评价与 meta 分析两个名词常被混用，但系统评价不一定都包括有 meta 分析过程，而 meta 分析也不一定是系统评价。

（四）文献综述（review）

文献综述又称为叙述性文献综述（narrative review）或传统文献综述（traditional review），由作者根据特定的目的和需要或兴趣，围绕某一题目收集相关的医学文献，采用定性分析的方法，对论文的研究目的、方法、结果、结论和观点等进行分析和评价，结合自己的观点和临床经验加以阐述和评论，总结成文，可为某一领域或专业提供大量的新知识和新进展，以便读者在较短时间内了解某一专题的研究概况和发展方向，解决临床实践中遇到的问题。但这种传统的文献综述，往往受综述者主观思维以及某些选择及测量性偏倚的影响。故在接受或应用这类证据时，宜持谨慎态度。

二、为什么要进行系统评价

作为一种重要的科研方法，进行系统评价（systematic review），有其一定的必然性：

（一）应对信息时代的挑战

每年约有 200 万篇生物医学文献发表在 2 万多种生物医学杂志上，年增长率约为 6.7%。一个内科医师需要每天不间断地阅读 19 篇专业文献才能勉强掌握本学科的新进展、新研究结果，使得需要大量信息进行科学决策的临床医生、研究人员和卫生部门的决策者往往陷入难以驾驭的信息海洋之中。而系统评价采用严格的选择、评价方法，去粗取精、去伪存真，将真实、可靠而有临床应用价值的信息进行合成，可直接为各层次的决策者提供科学依据。

（二）及时转化和应用研究成果

由于疾病谱的变化，对多因素疾病如恶性肿瘤、心脑血管疾病和各种慢性疾病的治疗方法的评估，需要尽量开展大样本临床试验，特别是随机对照试验（RCT）。但实施大规模的 RCT 需要消耗大量的人力、财力和时间，往往超过一个单位的承受能力，可行性受一定的限制。而现有的临床研究虽然数量多，但多数样本量不够大，故单个试验的结果难以提供较为全面、准确和推广应用价值大的研究结果。

因此，将多个质量较高的同质临床试验结果应用系统评价方法进行合成，则可将其综合的有效措施，及时转化和应用于临床实践与决策。如采用累积性 meta 分析回顾性分析有关静脉链激酶治疗急性心肌梗死的临床试验，1973 年前发表的 8 个 RCTs（2432 例患者）的 meta 分析即证明静脉链激酶能有效降低 AMI 患者的总死亡率（$P=0.01$）；1978 年前发表的 25 个 RCTs

笔记

(34 542 例患者)的 meta 分析，$P=0.001$(包括 GISSI-1 和 ISIS-2)，到 1986 年，$P=0.0001$，但静脉链激酶直至 1987 年才在传统综述和教科书中推荐常规用于治疗急性心肌梗死。可见，临床上整整晚了 14 年才开始应用 meta 分析结果，这期间如早用，则可挽救多少急性心肌梗死患者的生命。所以，Murphy 等在 1994 年指出：1973 年以后的大型临床试验，如果无医德问题，也是多余的，且花费大量经费。

（三）提高统计效能

针对同一临床问题的研究非常多，但因疾病诊断标准、纳入研究对象的标准、测量结果方法、治疗措施和研究设计等的差异，结果可能不一致，甚至相互矛盾。如对可能早产的孕妇使用激素的例子，纳入的 7 个高质量临床试验，尽管只有两个试验结果有统计学意义，但对 7 个临床试验进行定量系统评价后，样本含量和统计效能得以增加，合并效应量变得有统计学意义，即肯定了糖皮质激素能有效降低早产儿的病死率。系统评价或 meta 分析在进行资料合成时，不是根据阴性或阳性研究的个数多少来决定哪种治疗措施有效，而是综合考虑了各个研究的样本量大小及其研究质量。

另外，系统评价可减少有关偏倚的影响，从而提高研究结果的真实性和准确性。

三、系统评价与叙述性文献综述的区别与联系

系统评价和叙述性文献综述均是对临床研究文献的分析和总结，目前多为回顾性、观察性的研究，也可为前瞻性系统评价。回顾性的系统评价因受到纳入的原始临床研究质量的制约，难以有效控制偏倚和混杂的影响。因此，确定一篇“综述”属于叙述性文献综述还是系统评价？其质量、价值如何？主要取决于是否采用科学的方法减少了偏倚、混杂因素的影响。

叙述性文献综述常常涉及某一问题的多个方面，如糖尿病的病理、病理生理、流行病学、诊断方法及预防、治疗、康复的措施，也可仅涉及某一方面的问题，如诊断、治疗等。而系统评价或 meta 分析的研究问题往往比较具体明确。因此，叙述性文献综述有助于广泛了解某一疾病的全貌，而系统评价则有助于深入了解某一具体疾病的某一具体方面。二者的主要区别见表 7-1。

表 7-1 叙述性文献综述与系统评价的区别

特征	叙述性文献综述	系统评价
研究的问题	涉及的范畴常较广泛	常集中于某一临床问题
原始文献来源	常未说明、不全面	明确，常为多渠道
检索方法	常未说明	有明确的检索策略
原始文献的选择	常未说明、有潜在偏倚	有明确的选择标准
原始文献的评价	评价方法不统一	有严格的评价方法
结果的合成	多采用定性方法	定性、定量分析均可，以定量方法居多
结论的推断	有时会遵循研究依据	大多遵循研究依据
结果的更新	未定期更新	定期加以更新

第二节 系统评价的方法

笔记

系统评价是一把双刃剑，一方面能够通过对多个有争议或甚至相互矛盾的小型临床研究采用严格、系统的方法进行评价、分析和合成，以解决争议或提出建议，指导临床实践、医疗决策以及明确今后的研究方向；另一方面，如果纳入的原始研究质量不高或进行系统评价 /meta 分析的方法不当，会产生不正确的信息，造成误导。因此，系统评价方法和步骤，决定了其结果和结

论的真实可靠性。

系统评价本身只不过是一种研究方法，并不仅限于对随机对照试验或仅对治疗性研究进行系统评价。根据研究领域划分，系统评价可应用于基础研究、临床研究、医学教育、方法学研究、政策研究等领域；根据研究问题不同，系统评价又可应用于病因、诊断、治疗、预后、卫生经济评价和定性研究（qualitative research）等方面；若根据纳入的原始研究类型不同（study design），还可分为基于临床试验（controlled trial）和基于观察性研究（observational studies）的系统评价，前者如随机对照试验和非随机对照试验的系统评价，后者如队列研究和病例对照研究的系统评价；若按照纳入原始研究的方向性，又分为前瞻性、回顾性和累积性系统评价；根据资料分析时是否采用统计学方法（meta 分析），可分为定性和定量的系统评价。目前，鉴于随机对照试验的系统评价在理论和方法学上较完善且论证强度较高，数量较多，已成为实践循证医学的重要证据来源。

尽管系统评价的种类繁多，但基本方法和步骤相似（图 7-1），仅在检索策略、文献质量评价、数据提取以及汇总分析等方面略有不同，本节将以 Cochrane 系统评价为例，简述其基本方法和步骤。

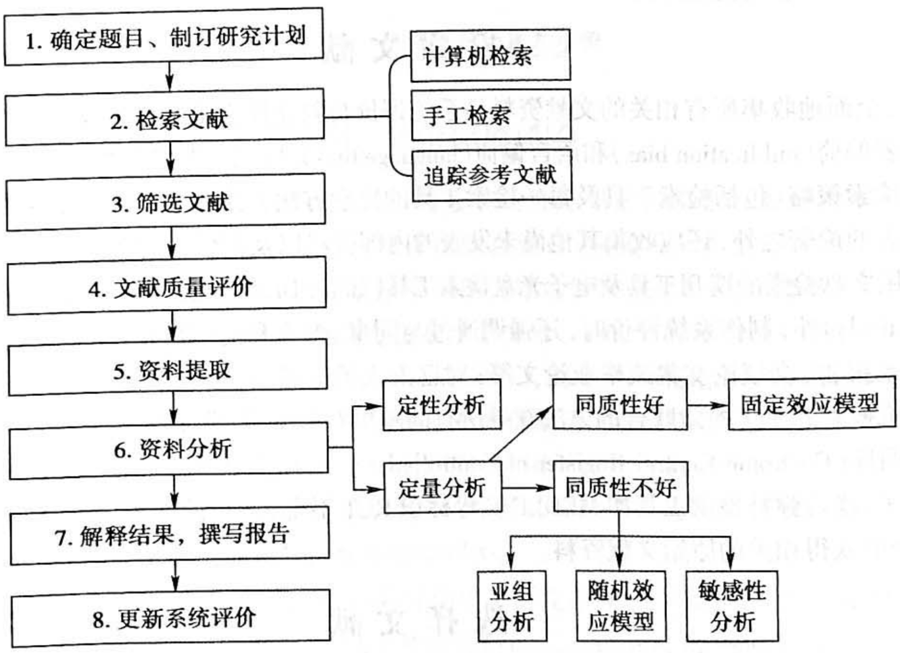


图 7-1 系统评价程序图示

一、确立题目、制定系统评价计划书

系统评价可为医疗保健措施的管理和应用提供重要的决策依据，特别适用于那些干预措施的利弊不明、靠单个临床研究结果难以确定或在临床应用过程中存在较大争议时使用。因此，系统评价的题目主要来源于临床医疗实践中那些不肯定、有争论的临床诊治问题。例如：在高危人群中服用小剂量的阿司匹林能否预防心脑血管病的发生？抗凝剂治疗能否预防缺血性心脏病伴心房纤颤患者继发心脏事件？急性胆囊炎患者，早期（发病后 7 天内）与延缓（入院治疗后 6 周）行腹腔镜胆囊切除术的疗效和安全性有无差别等？

为避免重复，首先应进行全面、系统的检索，了解针对同一临床问题的系统评价或 meta 分析是否已经存在或正在进行。如果有，其质量如何？是否已经过时？如果现有的系统评价或 meta 分析已过时或质量差，则可考虑进行更新或重新制作一个新的系统评价。

笔记

系统评价解决的问题很专一，涉及的研究对象、设计方案以及治疗措施需相似或相同。因此，在确立题目时，应围绕研究问题明确四个要素：①研究对象的类型：所患疾病类型及其诊断标准、研究人群的特征和场所；②研究的干预措施和进行比较的措施；③主要研究结果的类型，包括所有重要的结果（主要结果和次要结果）及严重的不良反应；④研究的设计方案。这些要素对指导查寻、筛选和评价各个临床研究，收集、分析数据及解释结果的应用价值等均十分重要，必须准确、清楚定义。

一旦确立了系统评价的题目，应着手制定计划书（protocol），内容包括系统评价的题目、背景资料、目的、检索文献的方法及策略、选择合格文献的标准、评价文献质量的方法、收集和分析数据的方案等。

原则上，系统评价研究的问题必须在制定计划书和收集文献前就已确定，这样可避免作者根据原始文献的数据信息和结果擅自更改系统评价的题目及内容，导致结论出现偏差。然而，由于多数系统评价是对现有文献资料的分析和总结，受原始文献及其质量的制约，如果事先不了解与题目相关的资料信息和内容，则难以确定一个好题目。因此，在进行系统评价的过程中如果要改变题目或评价的内容，必须明确回答原因及动机，并相应修改查寻文献和收集文献的方法。

二、检 索 文 献

系统、全面地收集所有相关的文献资料是系统评价与叙述性文献综述的重要区别之一。为了避免发表偏倚（publication bias）和语言偏倚（language bias），应围绕要解决的问题，按照计划书中制定的检索策略（包括检索工具及每一检索工具的检索方法），采用多种渠道和系统的检索方法。除发表的论著之外，还应收集其他尚未发表的内部资料以及多语种的相关资料。

除利用文献检索的期刊工具及电子光盘检索工具（如MEDLINE、Embase、SCISearch、Registers of clinical trials）外，制作系统评价时，还强调通过与同事、专家和药厂联系以获得未发表的文献资料如学术报告、会议论文集或毕业论文等；对已发表的文章，由Cochrane协作网的工作人员采用计算机检索和手工检索联合的方法查寻所有的随机对照试验，先后建立了Cochrane对照试验中心注册库（Cochrane Central Register of Controlled Trials，CENTRAL）和各专业评价小组对照试验注册库，既可弥补检索工具如MEDLINE等标识RCT不全、漏录的问题，也有助于系统评价者快速、全面获得相关的原始文献资料。

三、选 择 文 献

选择文献是指根据事先拟定的纳入和排除标准，从收集到的所有文献中检出能够回答研究问题的文献资料。因此，选择标准应根据确立的研究问题及构成研究问题的四要素即研究对象、干预措施、主要研究结果和研究的设计方案而制定。例如：拟探索静脉滴注硫酸镁能否降低急性心肌梗死患者的近期死亡率？围绕这一临床问题，如果确定研究对象为急性心肌梗死患者，不考虑梗死的部位、患者性别、年龄，干预措施为静脉使用硫酸镁与安慰剂比较，主要研究结果为35天内的病死率，设计方案为随机对照试验（RCT），则所选临床研究必须符合上述条件。而像口服硫酸镁或静脉滴注硫酸镁与其他药物进行比较、结果为心肌梗死35天以后的病死率或者非RCT的文献资料则不能纳入。

文献资料的选择应分三步进行（图7-2）：①初筛：根据检索出的引文信息如题目、摘要筛除明显不合格的文献，对肯定或不能肯定的文献应查出全文再进行筛选；②阅读全文：对可能合格的文献资料，应逐一阅读和分析全文后，再确认是否合格；③与作者联系：一旦被排除的文献将不再纳入，因此，如果文中提供的信息不全而不能确定者，或者有疑问和有分歧的文献应先纳入，通过与作者联系获得有关信息后再决定取舍。

笔记

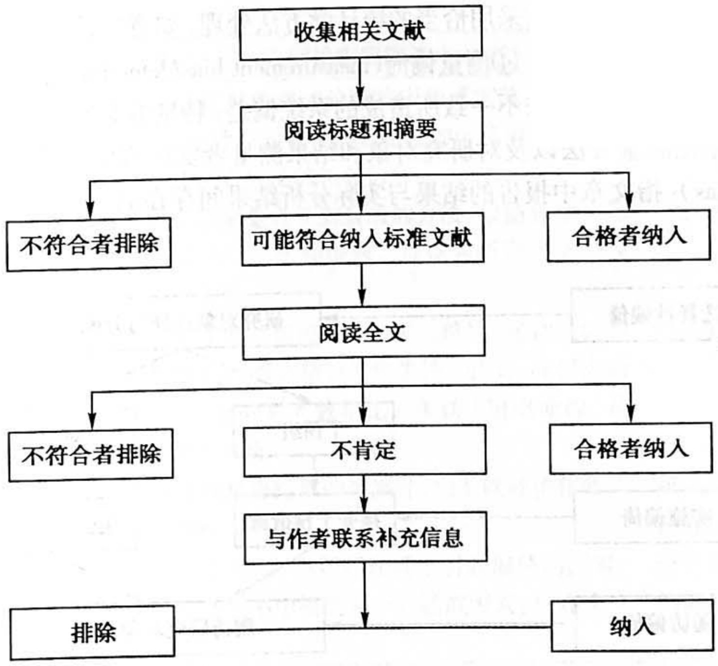


图7-2　选择文献的基本步骤

四、评价偏倚风险

原始研究的自身设计和实施质量如何，将影响研究结果的真实性。需对纳入研究的偏倚风险（risk of bias）进行评价，即评估单个临床试验在设计、实施和分析过程中防止或减少系统误差（或偏倚）和随机误差的程度。评价结果可作为原始研究文献的纳入阈值、用于解释不同文献结果差异的原因或用来计算敏感性分析和定量分析（meta分析）时的权重值。研究质量评价应包括两方面内容：①内在真实性（internal validity）：指单个研究结果接近真值的程度，即受各种偏倚因素如选择偏倚、实施偏倚、失访偏倚和测量偏倚的影响大小；②外部真实性（external validity，generalizability）：指研究结果是否可以推广应用于研究对象以外的其他人群，应结合结果的实用价值与推广应用的条件进行综合评定，外部真实性主要与研究对象的特征、研究措施的实施方法和条件以及结果的选择标准等密切相关。

研究质量评价（assessment of quality）是评估单个研究在设计、实施和分析过程中，防止或减少偏倚或系统误差的情况，也称为"方法学质量评价（assessment of methodological quality）"。一直以来，研究质量评价和偏倚风险评价被认为是等同的，但Cochrane系统评价手册认为偏倚和质量有区别。例如，某研究不可能对研究对象、干预措施的实施者或结果评价者采用盲法，若由此认定该研究的"低质量"，并不恰当。又如研究质量已达到了可能的最高水平，但并不是说该研究没有偏倚。另外，某些与研究质量相关的指标并不直接导致偏倚风险，如样本量估算、伦理审查和报告质量等。

偏倚是指导致研究结果偏离真值的现象，存在于临床试验中的不同阶段，如从选择和分配研究对象、实施干预措施、随访研究对象、测量和报告研究结果的每个阶段均可出现（图7-3）。因而，偏倚可分为以下五种：①选择偏倚（selection bias）：发生在选择和分配研究对象时，因随机方法不完善造成组间基线不可比，可夸大或缩小干预措施的疗效。采用真正的随机方法并对随机分配方案进行完善的隐藏，可避免此类偏倚的发生。②实施偏倚（performance bias）：发生在干预实施的过程中，除待比较的措施不同外，所提供的其他措施也不一样。采用标化治疗方案和盲法干预可避免实施偏倚。③随访偏倚（attrition bias）：指在试验随访过程中，试验组或对照组因退出、失访、违背治疗方案等造成人数或其他情况不一致而产生的系统偏差。对此，应尽量搜

笔记

集失访者的信息以及对失访人员采用恰当的统计学方法处理，如意向治疗分析(intention to treat analysis，ITT)等，可减少其影响。④测量偏倚(measurement bias/detection bias/ascertainment bias)：试验组和对照组的结果测量方法不一致所造成的系统偏差，特别容易在主观判断研究结果时出现。采用统一、标化测量方法以及对研究对象和结果测量者实施盲法，可避免其影响。⑤报告偏倚(reporting bias)：指文章中报告的结果与实际分析结果间存在的系统偏差，如有选择性地报告结果。

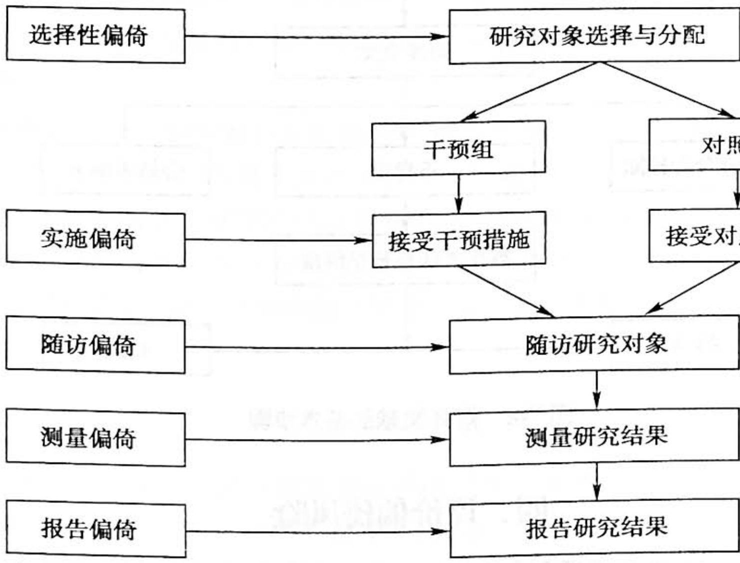


图 7-3　偏倚来源

目前，尽管评价文献质量的方法和工具较多，但缺乏"金标准"。评价工具可分清单或一栏表式(checklist，即有许多条目，但只打钩、不评分)和量表评分(scale，即有许多条目，每个条目均给予评分，进而给予相同或不同权重进行加权评分)。迄今至少有9种以上清单(checklist)和60余种量表(scale)用于评价随机对照试验的方法学质量，条目数从3个到57个不等，一般需要花10~45分钟完成评价。由于这些评价方法易受文献报告质量和文献评估者的主观因素影响，Cochrane手册5.0未推荐使用任何一种清单或量表，要求采用由Cochrane协作网的方法学家、编辑和系统评价员共同制定的新的"偏倚风险评估"工具(表7-2)，包括7个方面：①随机分配方法；②分配方案隐藏；③对研究对象、治疗方案实施者采用盲法；④对研究结果测量者采用盲法；⑤结果数据的完整性；⑥选择性报告研究结果；⑦其他偏倚来源。利用上述7条标准可逐一比对每一个纳入研究，并作出"偏倚风险低"、"偏倚风险高"和"偏倚风险不确定"的判断。其中，①、②、⑥、⑦条用于评估每一篇纳入研究的偏倚风险，其余3条则需针对每一篇纳入研究中的不同研究结果进行评估，强调同一研究中不同结果受偏倚影响程度不同。偏倚风险评价结果不仅采用文字和表格描述，还要求采用图示，更形象、直观反映偏倚情况。此评估工具对每一条的判断均有明确标准，减少了评估者主观因素影响，保证了评估结果具有更好的可靠性。

表 7-2　Cochrane协作网的偏倚风险评价工具

评价条目	评价内容描述	作者判断
①随机分配方法	详细描述产生随机分配序列的方法，以助于评估能否产生组间可比性	随机分配序列的产生是否正确
②分配方案隐藏	详细描述隐藏随机分配序列的方法，以助于判断干预措施分配情况是否可预知	分配方案隐藏是否完善

续表

评价条目	评价内容描述	作者判断
③对研究对象、治疗方案实施者采用盲法(针对每一研究结果评估)	描述对研究对象和试验措施实施者实施盲法的方法,以防止他们知道受试者接受的干预措施。提供判断盲法是否成功的相关信息	盲法是否完善
④对研究结果测量者采用盲法(针对每一研究结果评估)	描述对结果测量者实施盲法的方法,以防止他们知道受试者接受的干预措施。提供判断盲法是否成功的相关信息	盲法是否完善
⑤结果数据的完整性(针对每一研究结果评估)	报告每个主要结局指标的数据完整性,包括失访和退出的数据。明确是否报告失访 / 退出、每组人数(与随机入组的总人数相比)、失访 / 退出原因,是否采用 ITT 分析	结果数据是否完整
⑥选择性报告研究结果	描述选择性报告结果的可能性(由系统评价作者判断)及情况	研究报告是否提示无选择性报告结果
⑦其他偏倚来源	除以上 5 个方面,是否存在其他引起偏倚的因素?若事先在计划书中提到某个问题或因素,应在全文中作答	研究是否存在引起高度偏倚风险的其他因素

为避免选择文献和评价文献质量人员的偏倚,可考虑一篇文章由多人或盲法选评,也可采用专业与非专业人员相互配合、共同选评的办法,对选评文献过程中存在的分歧可通过共同讨论或请第三方协助解决。若多人选择文献时,应计算不同评价者间的一致性(Kappa 值)。此外,最好先进行预试验,以摸索经验,标化和统一选择、评价方法。

五、收集数据

根据制定的调查表和需要收集的内容,收录有关数据资料,其中包括:①一般资料:如评价的题目、评价者的姓名、原始文献编号和来源、评价的日期等;②研究特征:如研究的合格性、研究对象的特征和研究地点、文献的设计方案和质量、研究措施的具体内容和实施方法、有关偏倚防止措施、主要的试验结果等;③结果测量:如随访时间、失访和退出情况、分类资料应收集每组总人数及事件发生率、连续资料应收集每组研究人数、均数和标准差或标准误等。

所有的数据资料均要输入系统评价管理软件(Review Manager, RevMan),以进行文献结果的分析和报告。

六、分析资料和报告结果

对收集的资料,可采用定性或定量的方法进行分析,以获得相应的结果。

(一)定性分析(non-quantitative synthesis)

定性分析是采用描述的方法,将每个临床研究的特征按研究对象、干预措施、研究结果、偏倚风险和设计方法等进行总结并列成表格形式,以便纵览所有纳入研究的结果、方法及研究间的差异,同时帮助制订定量合成计划及其结果解释,因此,定性分析是定量分析前必不可少的步骤。

(二)定量分析(quantitative synthesis)

定量分析包括同质性检验(或异质性检验)、meta 分析和敏感性分析。

1. 异质性检验(heterogeneity) 定量系统评价或 meta 分析是将多个研究结果合并成一个效应值,但不同研究结果间不可避免存在差异,即异质性。异质性分三类:临床异质性(clinical heterogeneity),指不同研究中研究对象、干预措施和结果测量等存在的差异;方法学异质性(methodological heterogeneity),指试验设计和质量在不同研究中存在的差异;统计学异质性(statistical

heterogeneity），指不同研究中干预措施的效应值存在的差异，是临床异质性和方法学异质性共同作用的结果。异质性检验是指对不同原始研究结果之间的变异程度进行检验。如果检验结果有统计学意义，应解释其可能的原因并考虑是否进行结果合成。确定各研究结果是否同质有两种方法：一种是目测法，即通过绘制森林图观察各研究效应值的可信区间是否有重叠及其程度，如果可信区间差异太大，则放弃合成分析或采用随机效应模型。另一种方法是直接进行同质性检验（*Q* test，Chi-square test），在此基础上借助 I^2 定量估计异质性大小，如 0%~40% 表示异质性可能不重要，30%~60% 表示有中度异质性，50%~90% 表示有显著异质性，75%~100% 表示有很大异质性。

2. meta 分析　根据资料类型及评价目的选择效应量并对其进行定量合成分析。例如对分类变量，可选择优势比（odds ratio，*OR*）、相对危险度（relative risk，*RR*）、危险度差值（risk difference）和防止一例事件发生需要治疗同类患者的人数（number needed to treat，NNT）等作为效应量表达。对连续性变量，当结果测量采用相同度量衡单位时应选择均数差值（mean difference，*MD*），而当结果测量采用不同度量衡单位，如在不同研究中采用不同量表测试疼痛评分时，则应选择标化的均数差值（standardized mean difference，*SMD*）。用 meta 分析合成结果时，可选择固定效应模型（fixed effect model）或随机效应模型（random effect model），结果采用森林图（forest plot）表示。

3. 敏感性分析（sensitivity analysis）　指改变某些影响结果的重要因素如纳入标准、偏倚风险、失访情况、统计方法（固定效应或随机效应模型）和选择不同的效应量（优势比或相对危险度）等，以观察同质性和合成结果是否发生变化，从而判断合并结果的稳定性及其程度。

七、解释系统评价的结果和做结论

系统评价的目的是帮助患者、公众、医生、管理者和决策者进行卫生决策，旨在提供信息和辅助解释结果，而不是作出推荐意见。因此，清晰陈述研究结果、深入的讨论和明确的结论是系统评价的重要组成部分。解释系统评价和做结论必须基于研究的结果，内容应包括：

（一）系统评价证据的质量

Cochrane 协作网采用 GRADE（Grading of Recommendations Assessment，Development and Evaluation）分级和证据质量评价系统对系统评价的总体质量进行评估。该系统是 2004 年由包括 WHO 在内的 19 个国家和国际组织、67 名专家（包括临床指南专家、循证医学专家、各个标准的主要制定者及证据研究人员）共同成立的 GRADE 工作组循证制定出的、国际统一的证据质量分级和推荐强度标准，分别于 2008 年正式在 BMJ 发表 5 篇系列文章，2011 年再次完善、更新并在临床流行病学杂志系列发表 20 篇文章，为生产者和 GRADE 结果使用者提供了详尽指导。GRADE 质量评价系统将系统评价证据的质量分为高、中、低、极低 4 个等级，并根据纳入研究的总体偏倚风险、研究结果的一致性、证据的直接性、结果的精度以及是否存在发表偏倚等 5 个因素调整随机对照试验的质量级别；同时也可根据效应值大小、是否存在剂量 - 效应关系以及是否存在偏倚导致效应低估甚至无效的可能性等，适当提升观察性研究的质量级别。

（二）推广应用性

在确定系统评价结果的应用价值时，首先应考虑干预措施对患者的利弊关系，其次应考虑系统评价中的纳入研究，其研究对象是否与你的患者情况相似？是否存在生物学、社会文化背景、依从性、基础危险度、病情、价值观等方面的差异。

（三）解释统计分析结果

解释 meta 分析结果时，应同时考虑干预措施的利和弊，合并效应量及其 95% 可信区间。点估计值主要表示合并效应量的强度和方向，而可信区间则反映合并效应量的变动范围和精确性，二者结合可提供更全面的信息，有助于解释结果的临床价值。

（四）结论

系统评价的结论应包括对临床实践的意义和未来研究的价值两部分内容。在确定这两方面意义时，要考虑证据的质量、干预措施的利弊、患者的价值取向及卫生资源的利用，目的在于帮助医务工作者和决策者进行正确的选择和应用，为进一步的研究指明方向。

八、更新系统评价

系统评价的更新是指在系统评价发表以后，定期收集新的原始研究，按前述步骤重新进行分析、评价，以及时更新和补充新的信息，使系统评价日臻完善。

第三节　meta 分析

一、meta 分析的概述

meta 分析是一种定量合并方法。定量合并分析的想法最早是由统计学家 Karl Pearson 在 1904 年提出，但直到 1955 年才得以在临床研究中具体应用。G. V. Glass 于 1976 年在教育研究中正式将这类对文献进行综合研究的方法冠以“meta-analysis”。其中“meta”为希腊词，有“after”、“more comprehensive”、“secondary”之意。中文译名较多，有 meta 分析、荟萃分析、综合分析、元分析、二次分析等。其中以“meta 分析”最为常用。meta 分析通过对多个同类独立研究结果的汇总和合并分析，可以达到增大样本含量、提高检验效能的目的；同时也可提高效应量的估计精度；特别是当多个研究结果不一致或都无统计学意义时，用 meta 分析可得到更加接近真实情况的统计结果。

系统评价（或系统综述）中的汇总分析分为定性与定量分析两种。如果纳入的原始研究缺乏有效数据或者研究结果间差别过大，那么就无法进行定量评价，只能得到定性描述结果；相反，若条件允许，可考虑进行定量评价，即 meta 分析。meta 分析是将两个或多个相似研究结果进行定量综合分析的一类方法。广义上包括提出问题、检索相关研究文献、制定文献纳入和排除标准、描述基本信息、定量综合分析等一系列过程。狭义上，meta 分析则专指系统评价中的定量分析。

因此，在一个系统评价中可以选用某个结局指标进行一次 meta 分析，也可选用多个结局指标实施多个 meta 分析。事实上，由于纳入研究的质量、设计类型、资料类型以及方法学等限制，只有部分系统评价可以实现定量分析。

目前，有关系统评价及 meta 分析已经在临床研究和临床实践中得以普及与推广，特别是已被广泛应用于效应量较小或存在争议的治疗性研究（主要为 RCT）、预后研究、病因学研究等，并逐步推广到诊断试验的系统评价之中。

二、meta 分析的基本过程

meta 分析过程包括数据提取及汇总、异质性检验、模型选择、合并效应量估计及假设检验等。

（一）数据提取

数据是否准确可靠，尤为关键，它是 meta 分析的基础，否则即使再先进的统计学方法，也不能弥补数据本身的缺陷。所以在收集与提取数据时，应广开渠道，通过多途径收集，确保数据全面完整；同时，采取有效的质控措施，如多人同步提取数据，防止选择性偏倚；最后对数据资料自身的真实性也要进行严格评价，在此基础上，方可实施 meta 分析。

（二）数据类型及其效应量的表达

目前可用于 meta 分析的数据类型主要包括以下 5 类：①二分类变量资料，按照某种属性

分为互不相容的两类，如描述临床结局时，选用存活、死亡，复发或不复发等；②数值变量/连续性变量资料，如血压值、血糖、CD4/CD8等，往往有度量衡单位，且能够做到精确测量；③等级资料/有序多分类变量资料，即将某种属性分为多个类别，类与类间有程度或等级上差异。如疗效判定用痊愈、显效、有效、无效等表示。以上三类数据类型比较常见，此外还有计数数据或密度资料，即同一个体在一定观察时间内可发生多次不良事件，如心肌梗死、骨折、入院次数等；以及生存资料，同时观察两类数据，即是否发生不良事件以及发生不良事件的时间等。

不同数据类型决定了效应量的表达方式有所不同。效应量（effect size）常被定义为临床上有意义的值或改变量。当结局观察指标为二分类变量资料时，常用的效应量表达有相对危险度（relative risk，*RR*）、优势比（odds ratio，*OR*）、绝对危险度（absolute risk，*AR*）或*NNT*等；当结局观察指标为连续性变量资料、非罕发的计数数据、较多分类的等级资料时，效应量采用均数差值（mean difference，*MD*）或标准化均数差值（standardized mean difference，*SMD*）等表达方式。对于较少分类的等级资料或罕发的计数数据，可转化为二分类变量资料进行处理，并选用相应的效应量；对于类似发病密度的数据，可以使用risk ratio，也简写成*RR*。对于生存资料，效应量表达可用风险比（hazard ratio，*HR*）。

在此基础上，按照统一设计的数据提取表，系统收集所纳入研究的重要信息，如样本量、分析方法、主要结果变量、设计方案、发表年份、具体实施时间及地点、质量控制措施等。

（三）异质性检验（heterogeneity test）

meta分析之前，应进行异质性检验，并根据异质性检验结果，来决定是否估计合并效应量。异质性检验又称同质性检验，旨在检验多个原始研究结果间的一致性。异质性检验方法主要有Q检验法与图形目测法等。若Q检验有统计学意义，则表明存在统计学异质性（statistical heterogeneity），需要探讨异质性的来源并进行相应处理。异质性来源主要从两个方面考虑：一是临床异质性（clinical heterogeneity），如纳入研究在研究对象、干预措施、结局观察指标等存在差异；二是方法学异质性（methodological heterogeneity），如纳入了不同设计方案、不同方法学质量的原始研究等。

1. Q检验及I^2指数 Q检验的无效假设为：所有纳入研究的效应量均相同（即H_0：$\theta_1=\theta_2=\cdots=\theta_k$），$Q$统计量定义为：$Q=\sum w_i(\theta_i-\bar{\theta})^2$，进一步可表达为：$Q=\sum_{i=1}^{k} w_i\theta_i^2-\frac{(\sum w_i\theta_i)^2}{\sum w_i}$。

上式中w_i为第i个研究的权重值。θ_i为第i个研究的效应量，$\bar{\theta}$为合并效应量，$\bar{\theta}=\frac{\sum w_i\theta_i}{\sum w_i}$。

k为纳入的研究个数。Q服从于自由度为$k-1$的χ^2分布。若$Q>\chi^2_{(1-\alpha)}$，则$P<\alpha$，表明纳入研究间的效应量存在统计学异质性，可进一步计算异质指数$I^2=\frac{Q-(k-1)}{Q}\times 100\%$，用以定量描述异质程度。若$I^2$指数为0%~40%，表明异质性可以忽略不计；$I^2=30\%$~60%，表明存在一定程度的异质性；若$I^2=50\%$~90%，表明纳入研究的效应量存在较明显的异质性；当$I^2=75\%$~100%时，表明异质性明显，需探讨异质性来源，考虑进行亚组分析、meta回归等，甚至放弃meta分析。

例7-1 在一项有关丹参治疗急性期脑卒中的系统综述中，研究者纳入了5个随机临床试验，结局观察指标为神经功能缺损改善率，资料见表7-3。

以例7-1为例，Q检验法的计算过程见表7-4，其中$\theta_i=\ln OR_i$，$w_i=1/Var(\ln OR_i)$，$Var(\ln OR_i)=1/a_i+1/b_i+1/c_i+1/d_i$。

$Q=11.769-9.911^2/8.679=0.451$，$\upsilon=5-1=4$，$P>0.10$，异质性检验无统计学意义，还不能说明研究效应量间不同质，$I^2=0\%$。

需要注意的是Q检验法的检验效能较低，如在纳入研究数目较少的情况下，有时不能检测

笔记

表 7-3　丹参治疗急性期脑卒中的疗效观察

研究	干预组		对照组		合计(N_i)
	改善(a_i)	未改善(b_i)	改善(c_i)	未改善(d_i)	
Cao(1994)	48	1	45	2	96
Mao(2001)	33	3	25	9	70
Min(2004)	29	4	22	10	65
Pan(1992)	67	2	65	4	138
Zhang(2002)	36	4	29	11	80

表 7-4　Q 统计量计算过程(以 OR 为例)

研究	$\ln OR_i$	$Var(\ln OR_i)$	w_i	$w_i \times \ln OR_i$	$w_i \times \ln OR_i^2$
CAO(1994)	0.758	1.543	0.648	0.491	0.372
Mao(2001)	1.376	0.515	1.942	2.672	3.678
Min(2004)	1.193	0.43	2.326	2.773	3.307
Pan(1992)	0.723	0.78	1.282	0.928	0.671
Zhang(2002)	1.228	0.403	2.481	3.047	3.741
合计	1.142		8.679	9.911	11.769

出异质性，出现假阴性结果，可考虑提高检验水准，如 $\alpha=0.10$，以增大检验效能；相反，如果纳入研究过多，即使研究间结果是同质的，也可能出现 $P<\alpha$ 情况，即异质性检验有统计学意义。因此，对 Q 检验结果的解释要慎重，需要结合异质指数 I^2 以及森林图进行综合判断。

2. 图形法　此外，还有一些图形法用于展示异质性。如 Forest 图(森林图)、标准化 Z 分值图、Radial 图、L'Abbe 图等。其中通过目测森林图中的可信区间重叠程度，借以判断异质性最为常用。若可信区间大部分重叠，无明显异常值，一般可认定同质性较好。

(四) 合并效应量估计及其假设检验

在异质性检验的基础上，选用适当的方法进行分析。若异质性不明显，同时假定理论效应量为某一固定值，纳入研究效应量间的差异是由机遇造成的，可采用固定效应模型(fixed effect model)估计合并效应量；若存在一定程度异质性，且假定理论效应量不固定、服从于某种分布类型，如正态分布时，可用随机效应模型估计效应量；若异质性明显，可考虑亚组分析、meta 回归分析直至放弃汇总分析，只对结果进行简单描述。

以四格表资料为例，演示合并效应量估计及其假设检验过程。鉴于此类资料合并效应量的估计方法较多，诸如 Mantel-Haenszel 法、方差倒置法(inverse-variance methods)、Peto 法等。现仅以 Mantel-Haenszel 法为例，加以阐述(表 7-5)。

表 7-5　Mantel-Haenszel 法计算合并效应量的具体过程

研究	a_id_i/N_i	b_ic_i/N_i	OR_i
Cao(1994)	1.000	0.469	2.133
Mao(2001)	4.243	1.071	3.960
Min(2004)	4.462	1.354	3.295
Pan(1992)	1.942	0.942	2.062
Zhang(2002)	4.950	1.450	3.414
合计	16.596	5.286	3.14

笔记

1. 估计合并效应量及其95%可信区间 $OR_{MH}=\dfrac{\sum a_i d_i/N_i}{\sum b_i c_i/N_i}$，$OR_{MH}$ 对数的方差为 $Var(\ln OR_{MH})=\dfrac{\sum P_i R_i}{2(\sum R_i)^2}+\dfrac{\sum(P_i S_i+Q_i R_i)}{2(\sum R_i)(\sum S_i)}+\dfrac{\sum Q_i S_i}{2(\sum S_i)^2}$，其中 a_i，b_i，c_i，d_i 为四格表的实际频数，$R_i=\dfrac{a_i d_i}{N_i}$；$S_i=\dfrac{b_i c_i}{N_i}$；$P_i=\dfrac{a_i+d_i}{N_i}$；$Q_i=\dfrac{b_i+c_i}{N_i}$，则 OR_{MH} 的95%可信区间为：$\exp(\ln OR_{MH}\pm 1.96\sqrt{Var(\ln OR_{MH})})$

本例 $OR_{MH}=\dfrac{\sum a_i d_i/N_i}{\sum b_i c_i/N_i}=16.596/5.286=3.14$，$LnOR_{MH}$ 的方差为：0.115，则 OR_{MH} 值95%可信区间为1.62~6.10。

2. 合并效应量的假设检验：z 检验 $Z=\dfrac{\ln OR_{MH}}{\sqrt{Var(\ln OR_{MH})}}$，统计量 z 服从于 u 分布（外文文献常表示为 z 分布），用于检验合并效应量是否有统计学意义。

本例 $z=1.144/0.387=3.37$，则 $P<0.001$，表明合并效应量有统计学意义。

实际上，估计合并效应量以及进行异质性检验，可以借助一些现成分析软件来完成，方便快捷。这其中首推 RevMan 软件，图7-4则是利用该软件对例7-1资料的meta分析结果。

Study or Subgroup	Experimental Events	Experimental Total	Control Events	Control Total	Weight	Odds Ratio M-H, Fixed, 95% CI
CAO	48	49	45	47	8.9%	2.13 [0.19, 24.35]
Mao	33	36	25	34	20.3%	3.96 [0.97, 16.16]
Min	29	33	22	32	25.6%	3.30 [0.91, 11.91]
Pan	67	69	65	69	17.8%	2.06 [0.36, 11.64]
Zhang	36	40	29	40	27.4%	3.41 [0.98, 11.85]
Total（95% CI）		227		222	100.0%	3.14 [1.62, 6.10]
Total events	213		186			

Heterogeneity: Chi^2=0.45, df=4（P=0.98）; I^2=0%

Test for overall effect: Z=3.38（P=0.0007）

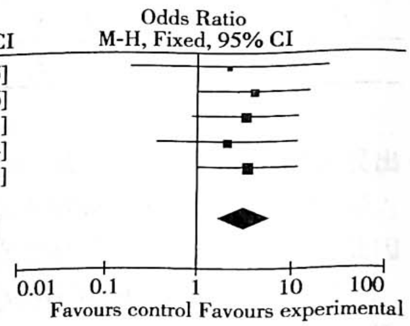


图7-4 基于RevMan软件的meta分析

图中“◆”为合并效应量图示，“3.14[1.62，6.10]”表示合并效应量及其95%可信区间；“$z=3.38$，$P=0.0007$”：表示假设检验中的统计量及其 P 值。“$Chi^2=0.45$，$df=4$，$P=0.98$”，表示异质性检验的 Q 值及 P 值，异质指数 $I^2=0\%$。

（五）固定效应模型与随机效应模型的选择

合并效应量的估计模型包括固定效应模型（fixed effect model）、随机效应模型（random effect model）以及最新提出的质量效应模型（quality effect model）等。模型的选择取决于异质性检验结果以及对效应量变异的理论假设。假如异质性检验无统计学意义且 $I^2<40\%$，并假设总体效应量为一个固定值时，可认为理论效应量是固定的，原始研究间的效应量即使有差别，也是由于抽样误差造成的，合并效应量估计可选用固定效应模型；当异质性检验有统计学意义（$P<0.10$）且 $I^2>50\%$，若假设合并效应量不固定并服从于服从某种分布（常假定为正态分布）时，考虑选用随机效应模型，计算合并效应量。随机效应模型因将研究间的变异因子 τ^2 作为校正权重，其结果比固定效应模型结果更稳健，但可信区间的精度会有所降低；若异质性明显（$I^2>75\%$），考虑meta回归、亚组分析，探讨异质性来源；若临床异质性过于明显，则应放弃进行meta分析，仅作定性描述。

1. 固定效应模型 上例数据类型为典型的二分类变量资料，选用固定效应模型的MH法估计得到了合并效应量。若遇到数值变量资料（连续性变量资料），且异质性检验无统计学意义时，同样可选用固定效应模型进行meta分析，具体过程与二分类变量资料相同，采用方差倒置法进行合并效应量估计。数值变量资料的效应量表达可以选择均数差值（mean difference，MD）

笔记

和标准化均数差值（standardized mean difference，SMD）。当纳入研究的结果变量均采用相同方式测量与表达时，效应量表达可使用均数差值（MD），因其带有自然单位，易于临床解释。而结果变量采用不同的度量衡单位或者效应量大小相差较大时，宜采用标准化均数差值（SMD），但需谨慎解释这类结果。现以 MD 作为效应量，阐述固定效应模型的估计过程。

（1）单个研究 MD 及其 95% 可信区间估计：MD 为均数差值，可表示为 $d_i=\overline{x_i^t}-\overline{x_i^c}$，其中，$\overline{x_i^t}$、$s^t_i$ 为第 i 个纳入研究中干预组的均数和标准差，$\overline{x_i^c}$，s^c_i 为第 i 个研究中对照组的均数和标准差，则 d_i 的标准误为 $(s^t_i\times s^t_i/n_i^t+s^c_i\times s^c_i/n_i^c)^{1/2}$。

（2）Q 检验及其 I^2 指数：$Q=\sum w_i d_i^2-\dfrac{(\sum w_i d_i)^2}{\sum w_i}$，进一步可计算异质指数 I^2。

（3）合并效应量（MD 合并值）及其 95% 可信区间：$d_{合并}=\dfrac{\sum w_i d_i}{\sum w_i}$，其 95% 可信区间为 $d_{合并}\pm 1.96\sqrt{\dfrac{1}{\sum w_i}}$，其中 w_i 为 d_i 标准误的平方。

（4）合并效应量的假设检验：$z=\dfrac{d_{合并}}{SE_{d(合并)}}$，$SE_{d(合并)}=\sqrt{\dfrac{1}{\sum w_i}}$。

例 7-2　在葛根素治疗急性脑梗死的系统评价中，研究者纳入分析了 4 个随机临床试验，均使用 NIHSS 量表测量了治疗前后神经功能缺损评分，具体结果见表 7-6。

表 7-6　神经功能缺损评分变化值 meta 分析过程

研究	葛根素组			对照组			d_i	w_i	d_iw_i	$d^2_iw_i$
	n_i^t	$\overline{x_i^t}$	s^t_i	n_i^c	$\overline{x_i^c}$	s^c_i				
Wu（2003）	30	14.36	18.36	30	19.11	13.41	−4.75	0.06	−0.28	1.31
Liu（2001）	50	14.53	17.96	50	20.18	12.41	−5.65	0.10	−0.59	3.35
Zhao（2004）	22	9.80	7.00	23	12.30	6.90	−2.5	0.23	−0.58	1.45
Cao（1999）	30	8.13	2.28	30	15.00	2.29	−6.87	2.87	−19.74	135.59
合计	132			133			−6.48	3.27	−21.19	141.70

本例合并 MD 为 −6.48，其 95% 可信区间为 −7.57~−5.40。合并 MD 假设检验 $z=11.72$，$P<0.000\,01$；异质性检验 $Q=4.39$，$P=0.22$，无统计学意义，$I^2=\dfrac{Q-(k-1)}{Q}\times 100\%=31\%$。鉴于本例的异质性检验结果无统计学意义且异质指数 <40%，考虑选用固定效应模型。同样，以该资料为例，可使用 RevMan 的固定效应模型分析，具体结果见图 7-5。

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
CAO	14.36	18.36	30	19.11	13.41	30	1.8%	−4.75 [−12.89, 3.39]
Liu	14.53	17.96	50	20.18	12.41	50	3.2%	−5.65 [−11.70, 0.40]
Wu	9.8	7	22	12.3	6.9	23	7.1%	−2.50 [−6.56, 1.56]
Zhao	8.13	2.28	30	15	2.29	30	87.9%	−6.87 [−8.03, −5.71]
Total (95% CI)			132			133	100.0%	−6.48 [−7.57, −5.40]

Heterogeneity: $Chi^2=4.37$, df=3 (P=0.22); $I^2=31\%$
Test for overall effect: Z = 11.72 (P<0.000 01)

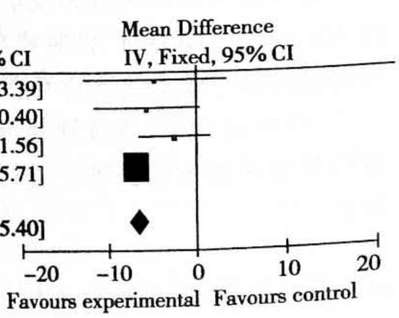


图 7-5　基于 RevMan 软件的 meta 分析

图中"◆"为合并效应量的图示，"−6.48[−7.57，−5.40]"表示合并效应量 MD $_{合并}$及其 95% 可信区间；"$z=11.72$，$P<0.000\,01$"：表示合并效应量的假设检验结果及其 P 值。"$Chi^2=4.37$，$df=3$，$P=0.22$"，表示异质性检验结果 Q 值及其 P 值。异质指数 $I^2=31\%$。由于 RevMan 在估计过程中

笔记

对MD及标准误进行了校正，结果与人工计算略有差别。

2. 随机效应模型 当异质性检验有统计学意义且假定真实效应量不固定但服从正态分布时，考虑选用随机效应模型（random effect model）估计合并效应量。随机效应模型就是在固定效应模型分析的基础上采用了DerSimonian-Laird校正。两类模型的区别在于加权的方式不同，固定效应模型以每个研究内方差的倒数作为权重，而随机效应模型是以研究内方差与研究间变异之和的倒数作为权重，调整的结果就是样本量较大研究的权重适当降低，而样本量较小研究的权重则适当增大。

（1）二分类变量资料的随机效应模型（random effect model）：该校正法是由DerSimonian和Laird于1986年提出，假设各原始研究的效应量不尽相同，以研究内方差及研究间变异之和的倒数为权重，并以此估计合并效应量。与固定效应模型相比，主要步骤相同，依次估计单个研究效应量、合并效应量及其95%可信区间，最后进行假设检验。唯一不同的是需事先计算研究间变异因子 τ^2。校正权重 $w_i^*=(w_i^{-1}+\tau^2)^{-1}$。其中 $\tau^2=\max\left[0,\dfrac{Q-(k-1)}{\sum w_i-(\sum w_i^2/\sum w_i)}\right]$，$Q$ 为异质性检验统计量，k 为纳入分析的研究个数。若 $Q<k-1$，$\tau^2=0$，若 $Q>k-1$，$\tau^2=\left[\dfrac{Q-(k-1)}{\sum w_i-(\sum w_i^2/\sum w_i)}\right]$。

仍以四格表资料为例，使用随机效应模型估计合并效应量及其95%可信区间：

$$OR_{合并}=\exp\left(\frac{\sum w_i^*\ln OR_i}{\sum w_i^*}\right)，95\%可信区间为\exp\left(\ln OR_{合并}\pm\frac{1.96}{\sqrt{\sum w_i^*}}\right)$$

以例7-1为例，$Q=0.451<4$，故 $\tau^2=0$，随机效应模型具体估计过程见表7-7。

表7-7　随机效应模型的分析结果

研究	OR_i	$\ln OR_i$	w_i	τ^2	w_i^*	$\ln OR_i\times w_i^*$
CAO（1994）	2.133	0.76	0.648	0	0.648	0.491
Mao（2001）	3.960	1.38	1.943	0	1.943	2.674
Min（2004）	3.295	1.19	2.326	0	2.326	2.774
Pan（1992）	2.062	0.72	1.282	0	1.282	0.927
Zhang（2002）	3.414	1.23	2.480	0	2.480	3.045
合计	3.14	1.14	8.679		8.679	9.911

其中 $OR_{合并}$ 为 exp（9.911/8.679）=3.133，其95%可信区间为1.62~6.10。由于固定效应模型以各研究内方差的倒数为权重，而D-L校正法则以研究内和研究间变异之和的倒数为权重，所以当异质性不明显时，两种模型估计结果完全相同；若存在较明显的异质性时，结果会有差异，利用随机效应模型估计的可信区间明显宽于固定效应模型的估计结果。

（2）数值变量（连续性变量）资料的随机效应模型：对于数值变量（连续性变量）资料，固定效应模型采用方差倒置法进行估计，而随机效应模型则是在此基础上进行DerSimonian-Laird校正。

校正权重 $w_i^*=(w_i^{-1}+\tau^2)^{-1}$，其中研究间变异因子 $\tau^2=\max\left[0,\dfrac{Q-(k-1)}{\sum w_i-(\sum w_i^2/\sum w_i)}\right]$，$w_i=1/v_{(di)}$，$Q$ 为异质性检验统计量，k 为纳入分析的研究个数。

MD合并值 $\bar{d}_{DL}=\dfrac{\sum w_i^* d_i}{\sum w_i^*}$，其假设检验统计量 $z=\dfrac{\bar{d}_{DL}}{1/\sqrt{w_i^*}}$，95%可信区间：$\bar{d}_{DL}\pm\dfrac{1.96}{\sqrt{\sum w_i^*}}$

以例7-2为例，使用随机效应模型重新分析，具体结果见表7-8。

笔记

表 7-8 神经功能缺损评分变化值 meta 分析过程

研究	葛根素组			对照组			d_i	w_i	τ^2	w_i^*
	n_i^t	$\overline{x_i^t}$	s_i^t	n_i^c	$\overline{x_i^c}$	s_i^c				
Wu(2003)	30	14.36	18.36	30	19.11	13.41	−4.75	0.06	1.917	0.05
Liu(2001)	50	14.53	17.96	50	20.18	12.41	−5.65	0.10	1.917	0.09
Zhao(2004)	22	9.80	7.00	23	12.30	6.90	−2.5	0.23	1.917	0.16
Cao(1999)	30	8.13	2.28	30	15.00	2.29	−6.87	2.87	1.917	0.44
合计	132			133			−5.63	3.27		0.74

本例 $Q>k-1$，$\tau^2=(4.39-3)/(3.27-8.322/3.27)=1.917$，合并均数差值：−5.63；合并效应量假设检验：$z=4.87$，$P<0.001$，说明有统计学意义；其 95% 可信区间为 −7.90~−3.37。同样可以利用 RevMan 完成上述分析，相对于固定效应模型的结果(−7.57~−5.40)，显然随机效应模型的区间范围变宽。

随机效应模型可以允许研究间效应量存在一定的异质性，并试图用特定的权重系数来解释这些异质性，所有纳入研究均使用相同的权重系数进行权重校正。但备受争议的是这些权重系数也许并无实际意义，所有研究都采用同一权重系数的做法也值得商榷。为此，Doi 和 Thalib 两位研究者提出的质量效应模型(quality effect model)可以较好地解决上述问题，他们认为异质性大多与原始研究的方法学质量有关，提出了偏倚风险概率(Qi)，以其作为权重校正系数，Qi 取值范围为 0~1，若 Qi 为 0，表示该原始研究质量高，无偏倚发生风险。目前已有一款免费软件 MetaXL，可以利用质量效应模型估计合并效应量。

三、发表性偏倚的识别与分析

meta 分析为一种二次研究方法，即基于原始研究结果进行二次分析。纳入的原始研究是否全面无偏，将直接影响 meta 分析结果是否真实可靠。在可能影响 meta 分析结果真实性的偏倚中，发表性偏倚的影响程度较大且较难控制，因而倍受关注。发表性偏倚可使 meta 分析过分夸大治疗效应量或危险因素的关联强度，导致临床个体治疗与卫生决策的失误。

发表性偏倚通常是指有统计学意义的研究结果比无统计学意义的研究更容易投稿和被发表，由此而产生的偏倚。对于无统计学意义的研究，研究者可能认为意义不大，不发表或推迟发表；作为杂志编辑则更有可能对这类论文退稿。因为存在发表性偏倚，即使具备周密的检索策略和手段(如与研究者个人联系)，也不可能完全地纳入所有相关研究。发表性偏倚的类型较多，常见的有：①当完成的临床试验得到阴性结果时，因研究者缺乏信心向国际知名的医学杂志投稿，而转投地方性杂志。②如非英语国家研究者，可能发表于本国的地方性杂志；但当得到阳性结果时，则作者更愿意在国际性杂志上用英文发表，这种发表性偏倚被称为语言性偏倚。③另外还有一些论文不能发表的原因，如博士、硕士读完学位而离开原来研究单位而未能发表。④或者一些研究结果可能违背了经费提供方(如药企)的利益，被迫搁浅不能发表。⑤出现发表性偏倚的另一种极端情况是，一些作者为提高知名度而一稿多投，或者作为多中心研究的参研单位，同时报告各自部分结果，造成多重发表性偏倚。

现有三类比较简单的分析方法：即漏斗图法、剪补法以及公式法可以用来正确识别与处理发表性偏倚。其中以漏斗图法最为常用，它是基于样本含量(或效应量标准误的倒数)与效应量(或效应量对数)所绘制的散点图。效应量可用 *RR*, *OR*, *RD* 或者 *RR*、*OR* 的对数值等。漏斗图的前提假设是效应量估计值的精度随着样本量的增大而提高，其变化范围也随精度的增加而逐渐变窄，最后趋近于点状，其形状类似一个对称倒置的漏斗，故称为漏斗图。即样本量小的研究，数量多、精度低，分布在漏斗图的底部呈左右对称排列；样本量大的研究，精度高，分布在

笔记

漏斗图的顶部，且向中间（合并效应量）集中。利用漏斗图可以直接观察原始研究的效应量估计值是否与其样本含量有关。当存在发表性偏倚时，漏斗图往往呈现不对称的偏态分布（图7-6）。但绘制漏斗图，需要纳入较多的研究个数，原则上要求5个以上才能进行。

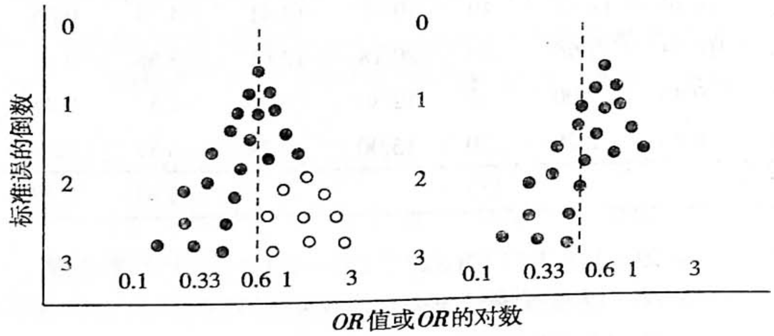


图7-6　漏斗示意图

图7-6所示假设为漏斗图的两种情况，左图中所有研究围绕中心线对称排列，表明没有发表性偏倚，图中空心散点代表结果无效的小样本研究，小样本研究估计的效应量变异较大，出现效应量极端值机会要多于大样本研究；右图，呈不对称分布，表示存在发表性偏倚，所缺失部分恰恰为结果无统计学意义的小样本研究。基于例7-1资料，利用RevMan5绘制漏斗图（图7-7）。

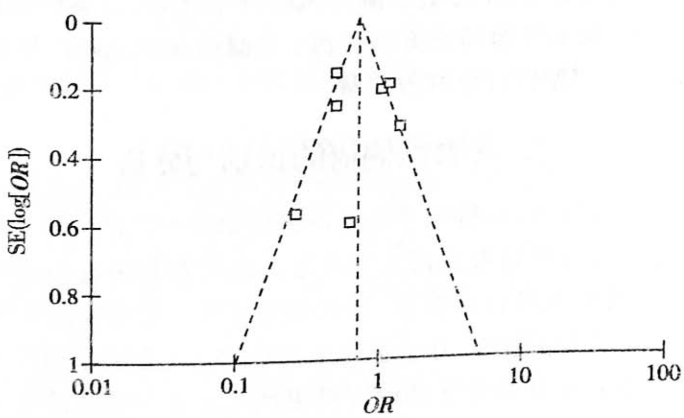


图7-7　例7-1漏斗图

图7-7显示所有研究围绕中心线对称排列，表明发表性偏倚不明显，对meta分析结果的影响可以忽略。

除漏斗图法外，也可以进行Egger回归、Begger分析以及计算失效安全数（fail-safe number）等，用以评估发表性偏倚。前者分析需要借助一些特定的软件，如Stata等。

四、meta回归及其他一些方法学进展

在临床研究中，即使研究目的完全相同，总会或多或少地存在一些差别。如在药物生产厂家、剂量、研究对象特征、病情轻重、测量时间、随访时间等方面有所不同，这些都是异质性的潜在来源。若这些因素能够被准确测量，可以选用meta回归模型，估计合并效应量。

$$\theta_i=\beta_0+\beta_1\times X_1+\cdots+\beta_p\times X_P+e_i$$

其中β_0，为固定效应量。若无混杂的影响，$\beta_1,\cdots,\beta_p=0$，则meta回归模型可简化为固定效应模型。meta回归模型可适用于RCT及病例对照研究等研究类型，也可用于敏感性分析。但meta回归容易产生聚集性偏倚，特别是当资料不齐或纳入分析的研究数目较少时，如小于10个时，不宜进行meta回归分析。

笔记

尽管上述回归模型中考虑了一些混杂因素，仍不能完全解释研究间的变异，可进一步在模型中加入随机效应项，那么该模型就成为混合效应模型。

$$\theta_i=\beta_0+\beta_1\times X_1+\cdots+\beta_p\times X_P+u_i+e_i$$

其中 u_i 为随机效应项。混合效应模型的参数估计可采用加权最小二乘法或极大似然估计法，可以用来解释已知的异质性来源。但存在两大缺点：一是如果研究数目较少，如小于 10 个，则不能建立混合效应模型；二是不能进行剂量反应回归分析等。

除 meta 回归模型与混合效应模型外，其他相关的方法还有：累积 meta 分析、迭代随机效应模型、多水平 meta 模型以及贝叶斯 meta 分析等。

五、慎重应用与评价 meta 分析的结果

（一）异质性检验与处理

若研究间有足够的同质性，选用合适的模型（如固定效应模型或随机效应模型，两者均可）估计合并效应量；若存在异质性且来源已知，采用 meta 回归模型或亚组分析，估计合并效应量。若异质性检验有统计学意义但异质性来源未知，当假设研究间效应量虽不固定，但服从正态分布时，应选择随机效应模型的估计结果；若异质性过大，应放弃 meta 分析，只对结果作一般性描述。

（二）考察 meta 分析结果的稳健性

考察 meta 分析结果的稳健性，常常采用敏感性分析。敏感性分析（sensitivity analysis）即通过改变纳入标准（特别是那些尚有争议的研究）、排除低质量的研究、或采用不同统计方法 / 模型分析同一组资料，观察 meta 分析结果的变化情况，借以考察结果的稳定性如何。如在排除某个低质量研究后，重新估计合并效应量，并与未排除前的 meta 分析结果进行比较，探讨该研究对合并效应量影响程度及结果稳定性。若排除后的结果未发生大的变化，说明敏感性低，结果较为稳健可信；相反，若排除后得到差别较大甚至截然相反结论，说明敏感性高，结果的稳健性差，在解释结果和下结论时应非常慎重，提示存在与干预措施效果相关的、重要的、潜在的偏倚因素，需进一步明确争议的来源。

（三）meta 分析结果的适用性

合并效应量实际上是多个原始研究效应量的加权平均值，因此 meta 分析的结果在推广应用时，应注意个体对象的特征及生物学或文化变异、干预场所、干预措施及依从性、有无辅助治疗等。不宜推荐没有 meta 分析证据支持的建议。在无肯定性结论时，应注意区别两种情况，是证据不充分而无定论，还是有证据表明确实无效。

（四）meta 分析结果的时效性

同系统评价（综述）一样，meta 分析的结论同样不是一成不变的，它只是对现有资料综合分析的结果，随着新的研究资料的不断纳入，其结论应加以更新。

第四节　系统评价的评价原则

近年来，系统评价或 meta 分析的数量明显增多，方法日趋复杂，对临床医务工作者和卫生决策者也产生了重要影响。但一篇系统评价或 meta 分析，并不表示其结论的绝对真实、可靠。有研究者从与 meta 分析质量有关的 6 个方面（研究设计、不同研究的可合成性、偏倚的控制、统计分析方法、敏感性分析、应用性）对 86 篇基于随机对照试验的 meta 分析进行了评价，结果发现仅 28% 的 meta 分析合格。因此，读者在阅读或应用系统评价或 meta 分析的结论指导临床实践前，必须对其方法和每一个步骤进行严格评价以确定系统评价的结论是否真实、可信，否则有可能被误导。评价治疗性研究系统评价的基本原则可分为三方面：

笔记

一、系统评价的结果是否真实

(一) 是否根据随机对照试验进行的系统评价

作为评价干预措施疗效"标准设计方案"的随机对照试验，如能很好地控制各种偏倚因素的影响，由此产生同质性好的系统评价被认为是论证强度最高的研究证据。而根据非同质RCTs及非随机对照试验进行的系统评价易受偏倚因素的影响，其系统评价的论证强度必然降低。

(二) 是否采用广泛和详细的检索策略检索相关文献

从作者报告的文献检索方法中可明确收集的文献是否全面。由于标识不完整，一般的文献检索数据库如MEDLINE仅能检出库中收录随机对照试验的50%，而发表偏倚即阳性结果的文章更易发表的现象可能导致系统评价出现假阳性结果。因此，全面的文献检索应包括手检相关杂志、检索会议论文集、学位论文、厂家数据库和与已发表文献作者联系。此外，如果文献检索时限制语种，也可能影响系统评价结论。目前，多数杂志均要求系统评价作者按照PRISMA声明规范报告系统评价和meta分析全文，其中包含检索流程图，要求详细陈述检索结果和筛选流程，有助于读者判断检索的完整性和筛选的合理性。收集的文献越系统、全面，则结论受发表偏倚的影响就越小，可信度就越大。

(三) 是否评估纳入的单个研究的真实性

由于系统评价多为对原始文献资料的再分析和总结，因此，除了进行系统评价的方法要严格外，原始文献的质量非常重要。所以文中应详细描述评价单个研究文献质量的方法，最好为多人独立评价并有良好的一致性。

(四) 是否采用单个病例资料(或每个研究的合成结果)进行meta分析

采用单个病例资料进行的meta分析被认为是meta分析的标尺(yardstick)，具有根据各研究合成结果进行meta分析不具备的优势，如对来自不同研究的结果采用一致的定义和分界点，能从病人水平分析异质性并进行生存分析，用通常确定的亚组进行分析以检验和提出假设、通过与试验者联系可详细核查和反复校正资料，以明确随机化和随访资料的质量，通过现有病例记录系统(诸如死亡登记)更新随访信息等，将系统偏倚和机遇的影响减至最小程度。

二、系统评价的结果是否重要

(一) 不同研究的结果是否一致

如果纳入系统评价的每个临床研究，其治疗效果相似或至少疗效的方向一致，则由此合成的结果的可信度较高。因此，作者应对各个研究结果之间的相似性，即进行同质性检验。如果同质性检验有统计学差异，则应解释差异的原因并考虑合成结果是否恰当。

(二) 系统评价的疗效大小、疗效的精确性如何

在进行结果合成时，不能通过简单比较阳性研究结果和阴性研究结果的研究个数来确定系统评价的结论，而应该根据研究的质量和样本含量的大小对不同研究给予不同的权重值，并采用恰当的指标如优势比(odds ratio)、相对危险度(relative risk)、均数差值(mean difference)、防止一例事件发生需要治疗同类患者的例数(NNT)和统计方法如随机效应模型和固定效应模型等合成结果，同时计算相应的可信区间。

三、系统评价的结果是否能应用于我的患者

系统评价报告的结果是所有研究对象的"平均效应"，而你主管的患者不一定在研究对象的范围内，因此在考虑系统评价的结果能否应用于你主管的具体患者时应从四个方面进行考虑：

笔记

（一）我的患者是否与系统评价中的研究对象差异较大，导致结果不能应用于我的患者（Is our patient so different from those in the study that its results cannot apply）

可通过比较我的患者与系统评价中的研究对象在性别、年龄、合并症、疾病严重程度、病程、依从性、文化背景、社会因素、生物学及临床特征等方面的差异，并结合临床专业知识综合判断结果的推广应用性。

（二）系统评价中的干预措施在我的医院是否可行（Is the treatment feasible in our setting）

由于技术力量、设备条件、社会经济因素的限制，即使系统评价中的干预措施效果明显，有时在自己所在的医院却不能实施，难以应用于患者。

（三）自己的患者从治疗中获得的利弊如何（What are our patient's potential benefits and harms from the therapy）

任何临床决策必须权衡利弊和费用，只有利大于弊且费用合理时才有价值应用于患者。例如：告诉一名患者其患病的真实情况有助于早期治疗和获取患者的配合，但也增加了患者的心理负担，可能降低生存质量。

（四）对于治疗的疗效和不良反应，自己患者的价值观和选择如何（What are our patient's values and preferences for both the outcome we are trying to prevent and the side-effects we may cause）

循证医学强调，任何医疗决策的制定应结合个人的专业知识和经验、当前最佳的研究证据和患者的选择进行综合考虑，应以"病人"为中心而不是单纯治病，目前越来越强调患者参与医疗决策。

第五节 系统评价的应用

一、临床医疗的需要

随着循证医学的兴起，强调任何医疗决策的制定应遵循和应用科学研究结果，即应将个人的临床专业知识与现有的最好临床研究证据结合起来进行综合考虑，为每个患者作出最佳的诊治决策。除了高质量的原始论著外，系统评价的广泛应用正不断地改进和规范着医务工作者的医疗实践行为。例如：美国的政策研究所常应用系统评价的结果制定临床实践指南。有关为低血容量、烧伤和低蛋白血症患者常规补充白蛋白的系统评价发现，该种疗法导致英格兰和威尔士的死亡人数每年增加1000~3000人，这一结果引起了临床医师、科研人员和卫生决策者的极大关注，并呼吁禁止盲目使用白蛋白。英国伦敦St. George医院根据Cochrane系统评价结果改变了急性哮喘的治疗方案，预计一年可节约成千上万英镑。

二、科研工作的需要

临床科研要具有先进性、新颖性和临床价值，面对浩瀚的医学文献信息，研究人员必须查寻、阅读和评价相关领域的文献资料，掌握研究课题的历史、现状、发展趋势、存在问题、当前研究的热点与矛盾，提出选题、立题的依据，避免重复前人的工作，为研究工作提供信息资料和研究方向。目前，许多国家都非常重视高质量系统评价在临床科研中的价值。例如，英国国家医学研究会资助的临床试验，要求申请者回答是否已有相关的系统评价及其结论如何，如果没有相关系统评价或现有的系统评价没有明确结论而需要进一步研究，就会邀请系统评价的作者参与临床试验申请书的评审。

三、反映学科新动态

围绕专业发展的热点，纵览某一领域的最新文献资料，作好有关专题的系统评价，全面、深

笔记

入和集中地反映该领域目前的动态和趋势、存在的问题和发展的方向，以促进学科的发展，保证不断地吸收新知识、新营养而居于学科的前沿位置。

四、医学教育的需要

医学教育除了向医学生传授各种疾病的共同规律和特性方面的知识外，还应该及时传授某一疾病的最新进展以及新药物、新技术的发展情况。教科书由于出版周期长，常常难以反映最新动态。因此，医学教育者需要不断地阅读有关医学文献以更新知识。而系统评价是快速获取有关知识的途径之一。此外，撰写医学教科书也应吸纳系统评价证据。

另外，广大的基层医务工作者，由于工作繁忙、文献资源有限，为了知识的不断更新，可通过阅读有实用价值的、真实可靠的系统评价，作为学习新知识的继续教育资源。

五、卫生决策的需要

随着人口增长、年龄老化、新技术和新药物的应用、人类健康需求层次的提高，使有限卫生资源与无限增长的卫生需求之间的矛盾日益加剧，要求各级卫生管理人员制定卫生决策时应以科学、可靠的研究结果为依据，合理分配卫生资源，提高有限卫生资源的利用率。目前许多国家在制定卫生决策时均要以医学文献资料特别是系统评价为依据。例如：1990年，魁北克的卫生技术评估委员会发表了一篇有关使用造影剂后发生副作用的meta分析。报告明确指出，没有证据说明使用高渗造影剂比低渗造影剂更增加生命危险，仅严重副作用的发生率稍有增加。这一结果的公布使魁北克在1990—1992年间因使用低渗造影剂的医疗费用明显降低，净节约（除去处理严重副作用的费用）约1200万美元，即使保守估计，也可节约1000万美元左右。

总之，采用科学、严格的方法产生的系统评价能为临床医疗实践、医学教育、科研和卫生决策提供真实、可靠的信息，但在应用系统评价的结论时应该进行严格评价。

（李　静　康德英）

笔记

第八章　病因与危险因素的研究、评价与循证实践

第一节　病因与危险因素研究的意义

疾病的病因学研究一直都是医学探索的一个重要领域。基础、临床和预防医学各科专家都致力于疾病病因的研究，但不同学科研究病因的方法、手段，考虑问题的角度各有不同，甚至对病因概念的理解及用于判断病因的标准也不一致，因此，我们会在医学教科书中病因（etiology）、发病机制（pathogenesis）、致病机制（mechanisms）、危险因素（risk factor）等多个标题下看到有关病因的论述。如流行病学是从群体的角度，应用概率论和逻辑推理的方法探索疾病的病因和疾病发生的影响因素；而在临床实践中，任何一种干预措施，包括手术和药物，都可能引起严重程度不等的不良反应，研究不良反应，实质上也是确定因果关系，只是此处的"因"是指临床医生在疾病预防、诊治过程中采用的各种措施和方法，如诊断技术、手术和药物等。

病因学研究探索的是疾病发生的原因及相关因素间的相互效应和各因素对疾病发生、发展的影响。

在病因学研究（因果关系研究）中，必须强调的一点是：任何与结局（疾病）有关联的原因性暴露都必在结局发生之前。但发生在结局（疾病）之前的暴露未必就是病因。病因学研究是正确认识疾病发生和流行规律的基础，也是临床正确诊断、治疗和预防的科学基础，同时可为医学决策提供依据，从而获得疾病防治的最大效益。其研究意义表现在以下几个方面：

一、有助于疾病的临床诊断与治疗

通过病因学研究弄清病因，掌握其发病机制和转归，会帮助医师对患者进行正确的诊断和治疗。对疾病需要对因治疗，力求治本，从而获得好的疗效。

医师在临床诊疗过程中，面对患者的实际情况，总要探讨可能致病的直接病因和危险因素，有针对性地进行相关的体检和化验检查，得到必要的信息，进行临床综合分析，力求获得准确的病因诊断，从而有利于实施更有效的治疗。

如果通过现代的医学手段而未能明确病因的某种疾病，特别是对于人民健康危害性显著者，为了有效防治，务必从病因及发病的危险因素研究着手。例如：我国在20世纪50年代的某些地区于秋收季节，发生了钩端螺旋体病的流行，其中，有许多重症病例，发生肺大出血而死亡，当初怀疑"肺鼠疫"流行，采取了若干重大防疫措施，也造成了一些不必要的影响，后来弄清了该病系钩端螺旋体感染所致。当病因弄明白后，采取了有针对性地对患者的临床治疗及人群预防，从而大大地提高了钩端螺旋体病的防治能力，有力地保障了人民健康。

二、有助于疾病的预防

病因学研究结果，不仅有助于临床诊断与治疗，对于疾病的预防，同样有用。通过各种媒介，对社区人群进行健康教育，从而让社区人群了解疾病的病因和危险因素，提高他们对疾病病因和有效预防措施知识的认知程度，并有意识地改变自己的生活方式和行为，从而预防疾病的发生。

对于某些急性传染病，常为特异性的病原生物体致病，病因一旦明确，可以研究特异性高、免疫力强的疫苗，进行人群有效预防，如天花、麻疹、乙脑、脊髓灰质炎、肝炎等疫苗。同时，病原生物体的确定，又可研制特异度高的诊断试剂，进而又提高了疾病的诊断水平。

笔记

第二节 病因与危险因素研究的基本概念

人类对疾病病因的认识是一个循序渐进的过程，先后经历了单一病因论及多病因论两个阶段。

一、病因与危险因素

（一）疾病病因概念的发展

无论中国或外国，古代人常将疾病归因于鬼神、上帝及天意，平日靠求神拜佛或祈祷，以期消灾除病。公元前5世纪，由中国古代哲学思想中衍生出了阴阳五行学说——金、木、水、火、土，并用这一学说去解释人体生理现象和病理变化的规律，将疾病的发生与外环境的物质——金、木、水、火、土联系起来，从而产生了朴素唯物主义的病因观。公元前5世纪，在西方也出现了类似的观点，反映在Hippocrates所著的*On Airs*，*Waters and Places*之中。他们还认为，与水、土及风等有关，夏季有脾肿大、发热及腹泻，冬季常引起生痰及喉嘶哑。由此可见，在古代国内外医家们都注意到疾病与环境有密切关系。我国古代早就认为山间的"瘴气"是疟疾的原因。19世纪上半叶，Sydenham等人关于疾病的"瘴气学说（miasma theory）"在西方盛行。人们认为不洁的水和土壤里散发出来的污浊之气（瘴气）是使人发病的原因。故而强调应设法清除贫民窟和其他不卫生的地方的"瘴气"，以期减少疾病。

1．单一病因论 意大利Fracastoro（1479—1553）最早提出：特异的疾病与特异的"传染物"有关。拉开了特异病因论的开端。19世纪，随着疾病微生物理论的发展，Henle和他的学生Koch提出了推断独特的活微生物导致特异疾病的Henle-Koch原理，对推动人类病因学研究作出了巨大的贡献。该法则共有4条：①病原微生物必然存在于患病动物体内，但在健康动物体内不应出现；②从患病动物分离得到的病原微生物可以做纯培养；③将分离出的纯培养物人工接种易感动物，会出现该疾病所特有的症状；④从人工接种的被感染的动物体内可以再次分离出性状与原有病原微生物相同的纯培养物。Koch于1876年证明炭疽病符合这一原理，随后许多传染性疾病也得到了证实。该理论在病因学的发展史上是革命性的，为干预措施的实施铺平了道路。

尽管这个法则有些绝对化，但却反映出当时人们在病因认识方面有不小的进步，即病原微生物是传染病的必要病因，而且每种传染病都有自己的特异的病原微生物。这是关于疾病的生物学病因的重要萌芽。随着19世纪末期微生物学的出现和发展，人们形成了这样的概念，即每一种疾病必定是某一种特异的病原物所致。这就是疾病单一病因论的"特异病因学说"。但是它并不能解释复杂的病因效应。因为即使是单一的病因，也可以引起多种疾病，更不用说绝大多数疾病的发生与多种因素共同发生作用有关的情况，其忽视了社会和环境等因素对疾病的影响作用，因此，单一病因论有其明显的局限性。

2．多病因论 医学研究人员在长期的疾病防治实践中逐渐发现，疾病的产生并不完全依赖特异的病原物，还和环境及人体自身的多种因素有关。如在一些非传染性疾病的病因学研究中发现，一种疾病的发生往往是多种因素综合作用的结果，而且多种致病因素同时存在的危害性要比其中单一因素存在时严重得多。这是由于它们在人体内的致病效应上，会产生交互作用。即使是传染性疾病的发生，也与多种因素共同作用有关。如肺结核，其发生固然需要有结核分枝杆菌的存在，但个体的遗传易感性、营养状况、情绪状况、居住环境状况等均参与人体肺结核的发生过程。

例如：一个多致病因素致冠心病的多病因交互效应研究，以单纯的收缩压水平为基础，分别与血清胆固醇/高密度脂蛋白浓度比、糖尿病史、吸烟以及左心室肥厚（心电图）等多种危险

笔记

因素联合研究，观测5年以上，以研究这些因素对55岁男性发生冠心病的结果（发病率）（图8-1），可见凡具5项危险因素者发生冠心病的危险性为正常者的21.5倍，具2种危险因素者也达6~10倍。因此，在病因学研究中，了解多病因及其交互效应是非常重要的。

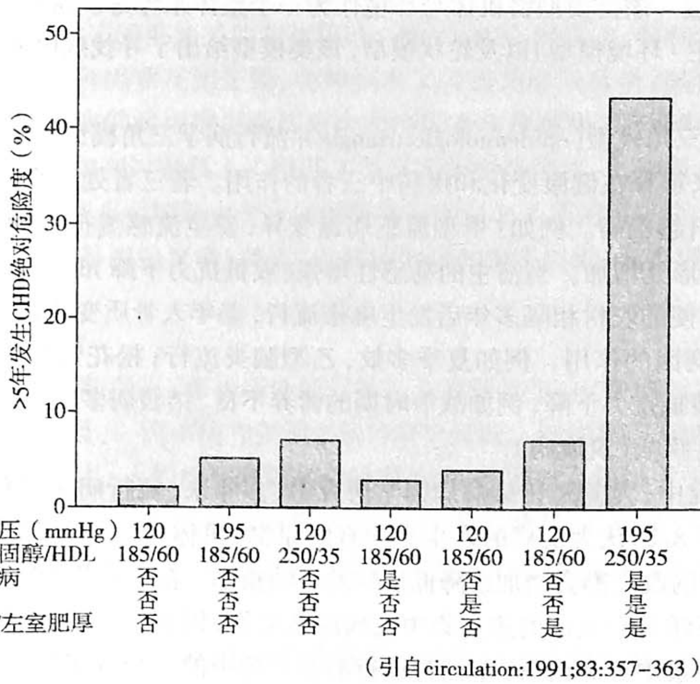


图8-1　冠心病的多病因交互效应图

3. 概率论因果观　人类对因果关系的认识一直处于发展之中。从古希腊学者亚里士多德（Aristotle）提出四原因说，到近代培根（Bacon）和休谟（Hume）提出决定论的因果观和经验论的因果观，走过了一个漫长的路程。然而，上述的因果论皆不能完美地解释生命现象中的因果关系。现代科学产生了概率论的因果观或称广义因果律（law of causation）。概率论的因果观认为原因就是使结果发生概率升高的事件或特征，即一定的原因可能导致一定的结果。该观点为解释生命科学中的因果关系奠定了理论基础。

Lilifield从概率论的因果观层面阐述了流行病学的病因概念，他在其所著的《流行病学基础》一书中给出的病因定义是："那些能使人群中发病概率升高的因素就可以被认为是病因，当其中的一个或多个因子不存在时，人群中疾病频率就会下降"。MacMahon也认为，流行病学的实际目的是发现能够预防疾病的联系，从这个目的出发，因果关联可以实用地定义为：事件或特征之间的一种关联，改变某一类别（X）的频率或特性，就会引起另一类别（Y）的频率或特性的改变，这样X就是Y的原因。因此，流行病学的病因观是符合概率论因果观的，流行病学层面的病因一般称之为危险因素（risk factor），这无疑体现了多病因论的思想，冲破了单病因论的束缚。概率论因果观的病因学定义不仅具有病因理论上的科学性和合理性，而且具有重要的公共卫生学意义。

例如：在20世纪60年代美国心脑血管病人的死亡率居高不下，经研究发现高血压为其主要危险因素，于是在全国开展了大规模高血压防治研究和人群防治运动。约10年后，高血压控制率大大提高，脑血管病死亡率大大降低，但冠心病死亡率下降不显著，于是又在全国开展了调节高脂血症的教育与防治，现在全民高脂血症有显著控制，心血管疾病死亡率呈现明显下降趋势，这些病因/危险因素研究符合概率论的因果观。

（二）病因模型

病因模型以简洁的概念关系图表达因果关系，这种在已有理论和经验基础上构建的概念关

系图，为我们提供了因果关系的思维框架。由于对因果关系有不同的理解或不同的侧重，所以研究者构建了多种类型的病因模型。目前具有代表性的危险因素作用模式有3类，即生态学模型（流行病学三角模式与轮状模式）、疾病因素模型和病因链与病因网络模型。

1. 生态学模型 该类模型将机体与环境作为一个整体来考虑。常见的有流行病学三角模型（又称病因-宿主-环境模型）以及轮状模型，该类模型给出了寻找病因的分类大框架，模型简明，整体性强。

（1）流行病学三角模型（epidemiologic triangle）：流行病学三角模型理论用病因、宿主和环境三者的平衡紊乱来解释在健康变化和疾病中三者的作用。若三者处于平衡状态，表现为健康；病因作用增强则引起疾病。例如：甲型流感病毒变异，发生流感流行；又如沙利度胺（反应停）投放市场，短肢畸形儿增加。当宿主的易感性增强（或抵抗力下降）时，同样可以导致疾病的发生。例如：交通不便的农村相隔多年后发生麻疹流行，老年人骨质变脆容易发生骨折。如果环境变化则可加重病因的作用。例如夏季多蚊，乙型脑炎流行；松花江污染，发生甲基汞中毒。环境变化还可促使抵抗力下降，例如战争时期的营养不良、结核病多发；或环境变化，因不适应环境的变化而发生疾病（筑波病）。

在病因学研究中，尤其是传染病病因学研究中，多年来，流行病学三角模型得到了广泛的认同。该模型（图8-2）认为疾病的发生必须有病原物、机体（宿主）及环境三个要素的协同作用，任何一个因素的改变都会增加或降低疾病发生的频率。在工业革命以前疾病谱呈现以传染病为主的特征，医学研究关注的焦点集中在病原微生物，因此，三角模型特别强调病原体的作用，将其从环境因素中分离出来，单列为流行病学三角中的一个重要部分。但是，随着社会的进步和科学的发展，即使对于传染性疾病，流行病学专家也更倾向于将病原微生物纳入到环境因素中，从而可以更全面地反映各因素之间的相互作用。更何况疾病谱已发生了很大变化，在以慢性非传染性疾病为主的今天，疾病的发生可能与社会经济、精神心理等多种因素有关，这些因素间的相互关系远非三角模型所能涵盖。

（2）轮状模型（wheel model）：1973年，Susser提出了疾病病因的轮状模型用以表示机体（宿主）与环境的关系，相对于流行病学三角模型，疾病病因的轮状模型更强调环境及其与机体间的密切关系。轮状模型（见彩图8-3）以宿主为轮轴，模型由外环和内环两部分组成，外环指的是环境，包括生物、理化和社会环境；内环指的是机体，包括人的自然特征（如年龄、性别）、营养状况、免疫力、内分泌水平和遗传等，其核心是遗传基因。外环生物环境包括病毒、细菌等各种病原微生物以及传播媒介和作为食物、制药原料的动植物；理化环境指的是气候、水、大气、土壤、光、辐射和各种化学物质如农药、杀虫剂和职业污染物等；社会环境包括社会经济水平、文化水平、政治制度、职业、居住条件、精神因素、个人行为方式等。轮状模型各部分的相对大小随不同疾病而有所变化。

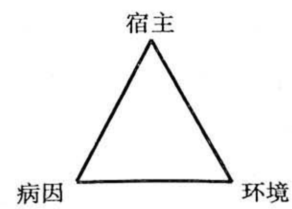


图8-2 流行病学三角模型

2. 疾病因素模型 该模型（图8-4）在病因分类上具有较强的可操作性和较强的实践指导意义。该模型将病因因素分为两个层次：外围的远因和致病机制的近因。外围的远因包括社会

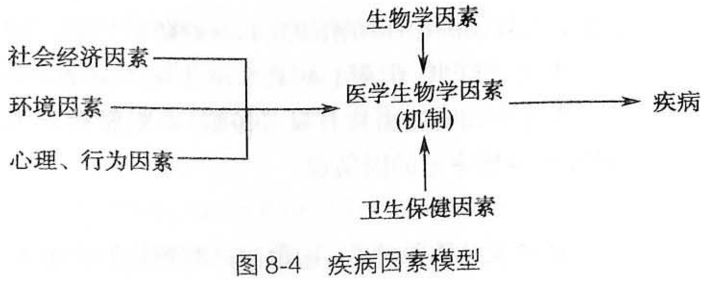


图8-4 疾病因素模型

笔记

经济、生物学、环境、心理、行为和卫生保健因素。基础或临床医学的病因主要是指致病机制的近因，临床流行病学病因学研究系以临床为基础，具有近因与远因相结合的特色。

3. 病因链与病因网络模型　1970年，MacMahon等提出了病因作用的网络模型，即疾病的病因因素按时间顺序连接起来构成病因链（chain of causation），多个病因链交错联接构成病因网（web of causation），它提供了因果关系的完整路径。MacMahon等指出，任何结果都不是由单一的孤立原因所致，而是各种因素互相交错，各种因素又各有其前因后果，其复杂程度远超过我们的想象限度。例如，肝癌的病因网可看成是由乙型肝炎病毒感染、饮用沟塘水、食用黄曲霉毒素污染的食物、遗传倾向和过量摄入乙醇等多条病因链交错构成，其中每条病因链又由多个环节连接构成，如饮用沟塘水可能是由于水中的藻类毒素引起了肝细胞恶变。

病因网络模型的优点是表示直观、具体，为病因阐述提供了依据，具有较强的可操作性，但其分析的技术难度较大。

（三）病因与危险因素

随着病因学研究的不断深入，多病因学说已被医学界所接受，已发现越来越多的疾病是多因性的，而且存在一病多因、一因多病和多因多病等复杂情况。按照病因与疾病间的作用方式、作用程度及传统哲学的观点，人们给病因以多种分类方法，本节仅介绍以下三种类型：

1. 必要病因与充分病因　按照传统的哲学观点，凡效应都有“必要条件”和“充分条件”之分，借助这种抽象的逻辑思维方式，我们可以认为任何疾病的发生都有相应的必要病因和充分病因。

（1）必要病因（necessary cause）：指有相应疾病发生以前，必定（概率为100%）有该因素存在。如果缺乏该因素即不会引起某疾病，则称该因素为该病的必要病因。如没有结核分枝杆菌感染就不会发生结核，因此，结核分枝杆菌是结核病的必要病因。绝大多数传染性疾病、职业病等都有一个比较明确的必要病因，而大多数慢性非传染性疾病目前尚未发现他们的必要病因。由于大多数慢性非传染性疾病是多因性的，因此，这类疾病可能不存在一个必要病因，或是必要病因仍待探究。

（2）充分病因（sufficient cause）：是指若有该病因存在，必定（概率是100%）导致某疾病发生。显然，充分病因即使针对传染性疾病也是非常少见的，因此，对充分病因的理解应注意三点：①对大多数疾病而言，充分病因的组成因素不是一个，而是一组。如上述的结核分枝杆菌感染仅是结核病的一个必要病因，而不是结核病的一个充分病因。因为大多数的结核分枝杆菌感染者可由于自身抵抗力的作用而不发生结核病，只有结核分枝杆菌感染结合机体特异性和非特异性抵抗力的降低，才能构成结核病的一个充分病因。②对于大多数疾病而言，其充分病因目前并未明了，一般只证实或初步证实充分病因中的个别或几个因素。③对于大多数慢性非传染性疾病来说，目前认为其充分病因不止一个，有的可能有多个充分病因，各充分病因的组成因素可能不同，因而这些疾病就可能没有必要病因。如肥胖（超重）是高血压的一个“病因”，但有的高血压病人并不超重，提示导致这部分病人发病的充分病因中可能不包括肥胖。

在日常生活中人们发现开启开关则电灯发光，于是便认为电灯开关是“因”，电灯发光是“果”，只要启动因，则必然获得果（灯亮）。这时此因对其果来说是必要而且充分的原因。但在电灯开关与电灯发亮的因果关系中，实际上有些重要因素被省略了。例如电线、灯泡、灯头、电流等。这些环节的任何一个都与开关同样重要，任何一个环节的缺少都会影响结果的产生。因此可以认为，任何结果的原因必然是由一组作用谐调的因素共同组成的，这就是充分病因。所以“充分病因”可以定义为：一组必然导致疾病的最低限度的状态或事件。这里的“最低限度”是指状态或事件的任何部分均是不可少的。这些组成充分病因的必不可少的部分称为成分病因（component causes）。

2. 直接病因与间接病因　基于病因链和病因网络模型，引起疾病的诸多因素有时可以连续

笔记

按顺次起作用，即病因 1 导致病因 2，最终导致疾病。可简要表示为：病因 1⟶病因 2⟶疾病。这里，病因 2 称为直接病因（direct causes），病因 1 称为间接病因（indirect causes）。直接病因是指只有该病因作用于人体才能够引起发病，对应于上述疾病因素模型中的近因。例如乙型肝炎病毒是乙型肝炎的直接病因；结核分枝杆菌是结核病的直接病因。间接病因实际上反映了引起疾病的阶段性或中间性过程，指可以促成和加速疾病发生的某些因素，其存在与疾病的发生呈间接关联，对应于疾病因素模型中的远因。例如：营养不良、居住条件差、机体免疫力低下、社会经济环境的恶化等都可能造成患病的易感性增加，这些因素被称之为间接病因。

再以世界卫生组织（WHO）2005 年提出的主要慢性病的病因链为例（图 8-5），高血压、高血糖、血脂异常和超重 / 肥胖这些因素就是所谓的直接病因或称之为“近端病因”，是医学界更为关注的主题。它们在病因链上距离疾病结局近，病因学意义相对明确，但是值得注意的是越靠近疾病结局近端的因素，涉及的人群面越窄，预防的机会越少。而从个体层面来看，不合理膳食及过多的能量摄入、体力活动少、吸烟则是上述直接病因共有的、最重要的、可以改变的危险因素，有效干预这三种危险因素可以预防 80% 的心血管疾病、2 型糖尿病和 40% 的肿瘤；再往病因链更远端看，还有“病因的原因”，即根本的社会经济、文化、政治和环境因素，称之为“健康社会决定因素”（social determinants of health，SDH）。此类远端影响因素作为间接病因，与疾病的因果机制可能不是那么明确与直接，但涉及的人群面广，预防机会大，通过改善这些因素对于降低总疾病负担的预防效率会很高。这些关于病因的认识和探讨势必会对疾病防治策略的调整产生深远的影响。

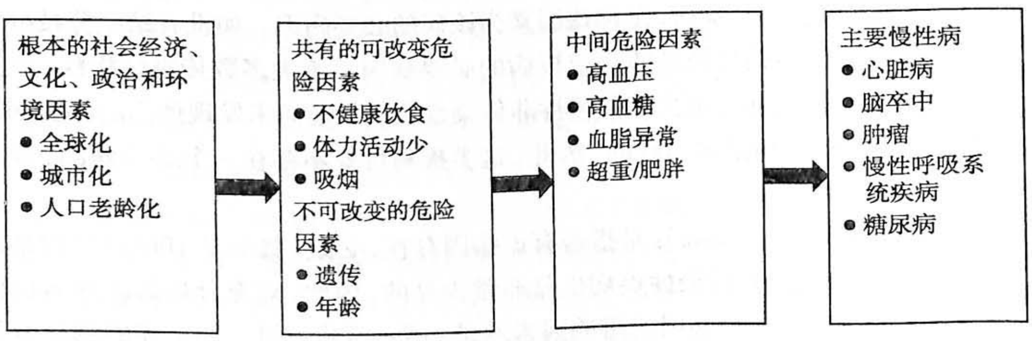


图 8-5 主要慢性病病因链

3. 危险因素 目前，慢性非传染性疾病已成为危害人类健康的主要疾病，慢性疾病由于发病比较隐匿，病程缓慢，病因复杂，从单一的患病个体去研究疾病病因会十分困难，因此，需要以相应群体作为研究对象对有关的发病因素进行宏观分析探讨，因而提出了危险因素这一概念。如前所述，一般将流行病学层次的病因称为危险因素，它是指疾病的发生与该因素有一定的因果关系，但是尚无可靠的证据能够证明该因素的致病效应，但是当消除该因素时，疾病发生的概率也随之下降。在病因学研究中，将这类与疾病发生有关的因素即称为危险因素。危险因素的概念无疑体现了概率论因果观。

二、寻找病因的指南清单

临床医师在进行病因学研究之前，一般须考虑选择对人类健康威胁较大、临床上存在医疗需要的疾病、病因不明的疾病进行研究，才更具有现实和科学意义。上述病因模型及病因分类为我们指出了寻找疾病病因的大致方向、类别或联系方式（病因网），但这些模型相对而言还比较抽象，不便于实际研究工作的开展，在病因学研究实践中，为了增加寻找病因的可操作性，我们可以从表 8-1 的寻找病因指南清单中得到启示。

笔记

表 8-1　寻找病因指南清单

一、宿主因素	1. 先天的	基因、染色体、性别
	2. 后天的	年龄、发育、营养状况、体格、行为类型、获得性免疫、既往史
二、环境因素	1. 生物的	病原体、感染动物、媒介昆虫、食入或接触的动植物
	2. 化学的	营养素、天然有毒动植物、化学药品、微量元素、重金属
	3. 物理的	电离辐射、噪声、振动、气象、地理(位置、地形、地质)
	4. 社会的	社会/人口(人口密度、居室、流动、都市化、交通、战争、灾害)、经济(收入、财产、景气)、家庭(构成、婚姻状况、家庭功能)、生活方式、饮食习惯、嗜好兴趣(烟、酒、茶、运动、消遣)、教育文化、医疗保健、职业(种类、场所、条件、福利、劳保设施)、政治、宗教、风俗

三、疾病自然史对病因学研究的意义

疾病自然史是指在不给予任何治疗和干预措施的情况下，疾病从发生、发展到结局的整个过程，疾病的自然史可以分为生物学反应期、亚临床期(或称临床前期，即症状出现前阶段)、临床期和疾病相关结局(如痊愈、残疾和死亡)四个时期。不同疾病其自然史差异很大，各期经历的时间长短，以及各个阶段的表现也有很大的差异。了解疾病的自然史对病因学研究有重要意义，主要表现在以下两个方面：

(一) 排除临床早期病例作为病因学的研究对象

为保证病因学研究结果的真实可靠，应避免将处于生物学反应期和亚临床期病例纳入研究对象(导致错分为“非病人群”)而造成选择性偏倚。这就要求在选择研究对象时，需要采用高敏感度的诊断试验，以发现处于生物学反应期和亚临床期病例，避免纳入可能的早期患者而影响研究质量。

(二) 确定有效的随访期、防止产生假阴性结果

对于若干慢性病发病后的自然病史，往往有一个较长的时间过程，其中生物学反应期和临床前期经历的时间也许较长。因此，在病因学的研究中，从其暴露于可能致病因素之后，到临床表现，需要一个时间上的效应过程。故应设计合适的致病效应观测期。如果观测期短，可能有些病例处于生物学反应期或临床前期而未能被确诊，于是就出现了“假阴性”结果而被漏诊，这就会影响研究结果的真实性。从疾病自然史角度出发，确定合理的随访期，并采用特异度和敏感度高的诊断试验方法，以助早期或不典型病例的确诊。这些在病因学研究中应有充分的认识。

第三节　疾病病因与危险因素研究的基本过程与方法

病因与危险因素作为重要的医学问题之一，可以根据对疾病的认识和掌握资料的程度分阶段进行研究。基本过程和可能采用的方法见图 8-6。

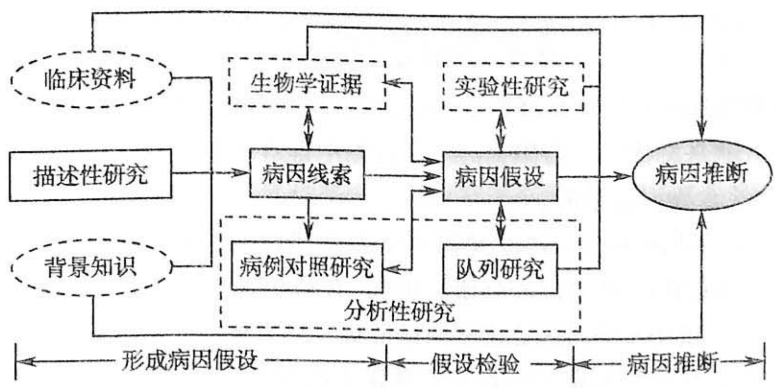


图 8-6　病因与危险因素研究的基本过程

一、提出病因假设

（一）假设的作用和特点

在病因探索研究中，形成病因假设，是关键的一步。假设是科学研究中一种广泛应用的方法，它是根据已知的科学原理和科学事实，对未知的自然现象及其规律性所给出的假定性说明或推测性解释。

在临床医疗实践或人群流行病学研究中，当发生了某种疾病，根据其基本背景知识、临床和流行病学史、临床体征、相关的实验室或特殊的检查资料，应用现代医学的知识和方法进行验证和推导，病因若尚不明，会为正确的诊断和有效防治带来很大的困难，这就迫使医学工作者必须对这种疾病的发病原因进行研究。例如：我国 2003 年春季在广州始发的 SARS，当初暴发流行，病情严重，病死率高，病因不明。对人群健康的威胁与危害十分严重，迫切需要弄清 SARS 的病因。

对于不明病因疾病进行病因学研究，首先必须面对被研究的疾病，依据它在人群中的分布特点、临床表现、病理损害的定位及其损害的程度，现有水平的各种化验、检查结果、综合分析，作出一系列的排除诊断，检索相关文献，在进行系统综合分析的基础上，作出可能的"假设"诊断，这就是要提出的病因假设。

例如：SARS 的暴发流行是通过呼吸道传染的。病因方面排除了细菌性感染、常见呼吸道病毒以及流感病毒感染。有人研究认为衣原体致病，其后被否定，于是提出了可能为一种毒力很强的未知病毒所致病因假设，通过国际协作研究，终于证实了 SARS 的病原体（病因）系变异型冠状病毒。

又如，我国华北地区大面积调查食管癌的流行特征后，通过资料分析与概括，发现该地食管癌死亡率较高的县、市，均集中在太行山脉南段三省（河北、河南、山西）交界处，并由此处向四周逐渐减低，与地势高低存在明显相关关系；而且发病率大体呈不规则的同心圆分布，最高和最低相差约 97 倍。研究者考虑到地理环境的可能重要作用，有人提出食管癌的水土病因学说。随后就有人根据这个假设，开展了地质化学方面的研究。

（二）形成病因假设的逻辑方法

前已提及事实依据和理论基础是建立假设的两个支撑点，但是，假设的形成不等于就事论事或事实与已有理论的混合。由事实和已有理论形成假设，必须经过一个较严密的逻辑思维过程。换句话说，假设形成的常用方法是逻辑方法。而假设形成过程中常用的逻辑方法主要是归纳演绎法。这种方法对于病因研究的因果假设有重要理论和实际指导价值。

1. 归纳法（Mill 准则）　J. S. Mill 是 19 世纪的哲学家，1856 年在他所著的《逻辑系统》一书中建立了数条准则，其中科学实验四法常被用于分析流行病学研究中形成假设、设计研究方案和进行病因推断。后人在科学实验四法的基础上将同异并用法单列，也即科学实验五法。

（1）求同法（method of agreement）：又称一致法或契合法或"异中求同法"，指对不同的事件或事物找出它们的共同点——共性。如在肝癌的病例对照研究中，肝癌病例组发现都有或相当部分有乙肝病毒感染标记；队列研究中，有乙型肝炎病毒持续感染者其肝癌发病率较高。提示乙肝病毒感染可能是肝癌的危险因素之一。食物中毒事件中，中毒患者大多有进食相同食物史，则可以提示由该食物引起中毒。

（2）求异法（method of difference）：又称差异法或"同中求异法"，指在相似的事件或事物之间找不同点（重要的差别）。还是以肝癌研究为例，在病例对照研究中，对照组多数不饮用沟塘水；在队列研究中，不饮用沟塘水的非暴露组的肝癌发病率低于饮用沟塘水的暴露组的发病率。提示两组与肝癌发生有关的差异之一是有无暴露于沟塘水。再如在非肝癌病例中发现都没有或相当部分没有乙肝病毒感染标记，表明乙肝病毒感染是肝癌的危险因素之一。

笔记

（3）同异并用法（joint method of agreement and difference）：求同法和求异法并用，相当于同一研究中设有比较组，控制干扰因素。

如宫颈癌的病因问题，据报道：性生活越是混乱的妇女发病率越高，早婚妇女的发病率又高于晚婚者，这是求同。与此相反，修女、尼姑与独身主义妇女很少患宫颈癌，这是求异。因此，有人提出性生活中的某因素可能与宫颈癌的发病有联系。随后的研究表明：宫颈癌可能与性交时的2型疱疹病毒感染有关。同异并用法是比较性研究（有对照组）设计的逻辑学基础。

（4）共变法（method of concomitant variation）：可以看成是求同法的特例。指研究因素的暴露程度不同时，疾病的频率也发生相应的变动，即在研究中注意发现疾病的患病率（有时是发病率）波动时有哪些因素在变动。共变法的应用有一定的条件，只有当有关（暴露）因素不是定性的，而是等级或定量的，并与事件（疾病）效应呈量变关系时，才可以应用共变法。

如在吸烟与肺癌的研究中，随着吸烟剂量（等级）的增加，肺癌的优势比（*OR*）或相对危险度（*RR*）也增加，即呈共变或剂量-反应关系，故支持吸烟是肺癌的病因的假设。

（5）剩余法（method of residues）：剩余法可以看成是求异法的特例，指当人们已知某复合结局事件（A、B、C）的有关暴露因素在特定的范围（a、b、c），通过事先的归纳又知道b说明B，c说明C，那么剩余的a必定说明A。剩余法就像算术中的减法，即在一组复杂的现象中把已知联系的现象去掉，探寻其他（剩余）现象的联系。如在肝癌的病因研究中，肝癌的发病率除了乙肝病毒感染和黄曲霉毒素能解释的部分，还有未能解释的部分，这部分或可归因于暴露因素范围内“剩余”的因素，如饮水中的藻类毒素。

需要注意的是：如果病因假设清单中没有包括真实的病因，Mill准则就不能提供任何帮助。遗憾的是，Mill准则对列出病因假设清单并不能提供指导，我们也无法知道要寻找的“那个”因素是否在清单中。

2. 假设演绎法　演绎是从一般到个别的推理。它是根据已知的规律来推论未知事物的方法，故又称类推法。假设形成后，通过假设演绎法同检验假设的分析性研究相衔接。

例如，我国原发性肝癌高发区主要分布在温暖、潮湿的东南沿海地区，在这些地区进行的大量描述性研究所获结果提示，乙型肝炎病毒感染、饮用沟塘水、食用被黄曲霉毒素污染的食物等因素的分布与原发性肝癌的分布相一致，从而为其后的分析和实验流行病学研究提供了线索，并形成了相应的工作假设。这一形成假设的过程衔接了描述性研究和分析性研究，其原理本质上是假设演绎法（hypothesis-deduction method）。其整个推论过程为：从假设演绎推出具体的证据，然后用观察或实验验证这个证据。如果证据成立，则支持假设的成立。从逻辑上看，反推是归纳的。从一个假设可推出多个具体证据，检验证实的具体证据越多，或证据的条件越多种多样，则归纳支持这个假设的概率越大。如果由假设演绎出来的具体证据不成立，并不能简单地否定假设，还需要考虑其他影响因素的存在。

以上的所谓归纳法以及假设演绎法，在病因学的假设中，可联系具体实际参考应用。

二、验证病因假设的要素

如上所述，科学的假设是在一定的经验材料和科学理论基础上经过逻辑思维加工提出来的，因此，同一个论题，由于事实依据不同，对科学理论的理解不同，思维方式不同等，常常会出现不同的假设。不仅不同的科研工作者对某一论题会提出不同的假设，甚至同一学者对同一论题，在不同时期提出的假设也会有差别甚至是很大的差别，这在医学对未知病因的研究，如像肿瘤、心脑血管病等病因多元性就是如此，由此来看，在科学上出现不同的假设，促进对一个问题的全面认识是很重要的一种正常的现象，也是科学发展兴旺发达的重要标志。

当然，科学上各种不同的假设有时受着主客观认识水平的限制，总有正确和错误之分，也有完善与不完善之别。但是假设的提出应持科学求实、严谨、创新的态度，力争使自己对所研

笔记

究问题假设，建立在较为可靠的科学理性基础之上，只有这样才有利于科学的发展。

（一）检验证实需要进一步深化

假设是否正确，需要通过实践（调查、观察和实验）来检验。检验的结果无非是证实或证伪。如果要证实自己研究的假设（如未知病因），那就应该全面进行研究，以探索在哪些条件下是符合的，在哪些条件下是不符合的，找出它的适用范围和局限性，并且深入地研究它的本质性内在联系，找出它的规律与机制，争取由假设上升到理性认识并指导自己的工作实践。例如SARS从临床流行病学研究深化到病原学的确定等，这些研究证实了“假设”，其成果又有效地指导了SARS防治实践。

（二）检验证伪应予区别对待

如果检验结果与假设不符，甚至完全相反，这表明假设是不够正确或是错误的。对于科学的发展来说，证实和证伪都具有重要意义。因为没有证实，不能肯定正确的假设；没有证伪，就不能否定错误的假设。因此，对于任何假设的验证结果，应持科学态度，进行具体分析，区别对待。一般而言，凡实验结果或观察到的现象与假设截然相反，或面对检验结果即使补充假设也无法自圆其说时，则应当放弃原有假设。若虽屡遭失败，但检验结果并不能否定假设的核心，或虽难以证实，但无直接否定假设的证据时，则不应随便放弃原假设，仍应从不同的角度或侧面，对其进行检验。

（三）检验假设的注意事项

检验假设最忌的是主观偏性。对于任何研究的假设无非有三种情况，一为通过验证，被确认为真理；二为受有关偏倚、混杂因素的影响得出虚假的因果联系；三为非科学的行为导致错误的因果联系。在病因学研究检验假设时，应避免或防止第二、三种情况的出现，这是要十分注意和坚持的，也是每位医学工作者务必养成尊重事实的思想作风与实事求是的工作态度。英国生理学家赫胥黎（Thomas Huxley）曾经说过：“我要做的是让我的愿望符合事实，而不是试图让事实与我的愿望调和。你们要像一个小学生那样坐在事实面前，准备放弃一切先入之见，恭恭敬敬地按照大自然指的路走，否则，就将一无所得。”只有与事实相符的假设，才有可能发展成为理论，进而促进科学的发展。

三、验证病因/危险因素假设的主要研究方法

对于病因/危险因素尚不清楚的疾病，如上所述，从背景知识或临床或流行病学的角度，总归会发现病因线索并提出病因的假设（假说），也许这种假设可能占有一些有说服力的证据，但都不能肯定或否定真正的病因，因此，必须通过科学的分析性或实验性的研究方可获得病因学的真实结论。

例如：原发性肝癌患者，从临床病例分析，多有乙型肝炎病毒感染——慢性肝炎的病史及其相应的临床特征和化验的证据。因而，临床医生从总结分析临床资料的角度，提出了肝癌的病因可能与乙型肝炎病毒感染、慢性肝炎病变有关。这个假说不经过科学的验证能肯定或否定吗？

又如：SARS暴发流行，临床发现一些病例病情危急，肺部炎变严重伴急性呼吸功能衰竭，病死率高，显属急性呼吸性感染致肺部重症炎变。经临床及实验室检验，排除了细菌性、流感病毒及常见呼吸道病毒的病原学诊断，临床提出了多为一种毒力很强的某一特殊病毒感染的假设（诊断），同样如不经过病原（病因）学的研究能肯定或否定这一假设吗？

因此，病因学研究往往需要多学科和多专业的协作和参与，才能获得病因学研究（病因假设的验证）的真正成功。

从病因学研究的角度，临床流行病学兼纳了临床医学和流行病学的两大学科优势，经过科学评价，列出以下常用研究方案及其证据的论证强度（科学性）（表8-2），供研究者应用时参考。

笔记

表 8-2 不同病因研究设计类型的论证强度

研究设计类型	性质	可行性	论证强度
随机对照试验	前瞻性	差	++++
队列研究	前瞻性	较好	+++
病例对照研究	回顾性	好	++
横断面研究	断面	好	+
叙述性研究	前瞻/回顾	好	±

(一) 病例对照研究

病例对照研究是分析流行病学中最基本的研究类型之一，性质上属回顾性研究。它是以确诊的患有某病的病人作为病例组，选择具有可比性的不患有该病的人群作为对照组。为排除混杂因素干扰，可以采用配比设计模式，比如用年龄与性别作为配比因素来排除年龄与性别可能的混杂作用，选择不患研究靶病的人，以 1∶1 或 1∶2 比例选择符合标准的对象作为对照组（参见本教材设计方案章节），然后进行病例组与对照组的比较分析，借助相应的统计学指标，探讨因果效应。

例如：HBsAg 与肝癌相关性的非配比设计的病例对照研究，结果见表 8-3。

表 8-3 HBsAg 与肝癌关系的病例对照研究

	病例组（肝癌）	对照组（非肝癌）	合计
HBsAg 阳性	35	8	43
HBsAg 阴性	15	42	57
合计	50	50	100

$OR=(35\times42)/(15\times8)=14$，结果显示 HBsAg 与肝癌发生有联系。

由于病例对照研究是从“果”到“因”的研究过程，且存在未知的混杂因素，论证效率相对较低。

(二) 队列研究

队列研究是重要的、可行性良好的病因/危险因素的研究方法，可以直接观察暴露于危险因素的不同人群的结局，具有较强论证力度。其是在特定人群中按照目前或过去是否暴露于待研究的危险因素，分为暴露与非暴露组，随访观察一段时间后，进而比较待研究疾病的发病率。由于为前瞻性研究，因此确定暴露与疾病的因果关系具有较强的论证力度。

例如：有一前瞻性队列研究，观测高胆固醇血症致冠心病的危险性，两个队列各为 1000 例 35 岁以上的观测对象，追踪 5 年，追踪率 90% 以上，结果高胆固醇血症队列发现冠心病 83 例；而正常血清胆固醇队列发现冠心病 23 例（表 8-4）。

表 8-4 高胆固醇血症致冠心病因果效应表

	冠心病		合计	发病率（%）
	+	−		
高胆固醇队列	83(a)	867(b)	950($a+b$)	8.7
正常胆固醇队列	23(c)	917(d)	940($c+d$)	2.5
合计	106($a+c$)	1784($b+d$)	1890(N)	

$$\text{相对危险度}(RR)=\frac{\dfrac{a}{(a+b)}}{\dfrac{c}{(c+d)}}=3.50\ (95\%\ CI:\ 3.48\sim3.52)$$

笔记

$$归因危险度(AR)=8.7\%-2.5\%=6.2\%$$

$$病因学分数(EF)=\frac{RR-1}{RR}=0.71$$

$NNH=\frac{1}{AR}=16$（表示高胆固醇血症5年中每16人就发生一例冠心病）。

结论：高胆固醇血症系冠心病发病的危险因素，其 $RR=3.5$（95% CI: 3.48~3.52）。

其中归因危险度（attributable risk，AR）也称特异危险度，又称率差（rate difference，RD）。临床试验中常把 AR 叫做绝对危险降低率（absolute risk reduction，ARR）。NNH（number needed to harm，可译为"害-需暴露人数"），具体概念详见本教材相关章节。

（三）实验流行病学研究

实验流行病学研究是一种可以施加干预因素，并且可控制研究条件的前瞻性的研究方法，比较给予不同干预因素后实验组与对照组的结局，借以判断干预因素对结局的影响，以佐证病因学的结论。这种实验性研究设计方案，视具体情况，可采用随机对照试验（RCT），也可采用非随机化的准实验研究设计模式。例如：我国大部分地区对新生儿和适龄儿童，采取了接种乙型病毒性肝炎疫苗以预防乙型病毒性肝炎的措施，通过 RCT 或准实验研究，追踪乙型病毒性肝炎的发病率，进而观测若干年，如肝癌发病率呈相应的下降，则可反证 HBsAg 与肝癌的病因的关系。实验流行病学的因果论证的强度较强。

（四）实验病因学研究

实验病因学从微观的角度去检验病因，借助生化实验、分子生物学实验、微生物学实验、动物实验等基础医学的研究，阐述病因的作用机制，为验证病因假设提供生物学证据。基础医学的微观研究对于认识疾病的本质以及对于疾病的进一步有效诊治具有重要价值。发病机制的研究成功能进一步肯定病因假设，加深对病因和致病过程的理解。但是同时应该注意到由于基础医学中模拟的人体内环境或者是动物实验均不可能代替人体本身，因此，实验结果外推到人这个过程需要谨慎。

四、病因学因果关联的统计学指标

用于病因学因果相关性分析的指标常用者如下：

1. **发病率**（incidence） 即暴露有关可疑病因或危险因素后，发病人数占其总体人数的百分比等。

2. **归因危险度**（attributable risk，AR 或绝对危险度 absolute risk，AR，又称率差 rate difference，RD） 即暴露可疑病因组人群和非暴露可疑病因组的人群，他们各自发病率的相减之绝对的差值。如A组发病率25%、B组为8%，则其归因危险度为25%－8%＝17%。

3. **相对危险度**（relative risk，RR） 即可疑病因接触组的发病率与对照组的发病率之比，常用来表示暴露与疾病联系的强度及其在病因学上的意义大小。

4. **相对危险增高度**（relative risk increase，RRI）或称病因学分数（etiologic fraction，EF） 是指暴露人群中发病或死亡归因于暴露的部分占全部发病或死亡的百分比，即绝对危险度与因素暴露组发病率的比值。

5. ***NNH***（number needed to harm，害-需暴露人数） 需要多少例接触致病因素后才出现一例发病，用 $1/AR$ 值表示。NNH 是临床和卫生决策十分有用且容易理解的指标。在疗效评价不良反应研究中，则可称之为害-需治人数，其对应于 NNT（number needed to treat，益-需治人数）。

上述这些评价因果之间关系的指标都可以计算各自的精确度（95% CI），至于因果相关性的强度和价值，则应结合具体的疾病情况和专业知识来确定。在此需要提醒的是，在不同的流行病学或循证医学教材或著作中各指标或有不同的称谓，读者应予注意。

笔记

五、防止偏倚干扰，慎重分析病因效应的关联

（一）虚假关联

在研究过程中，由于偏倚的干扰，会导致研究结果不真实，可以表现为夸大或掩盖了暴露因素与疾病之间的关联强度，甚至出现完全虚假的关联。例如，一项应用利血平治疗高血压的研究。结果发现利血平治疗有致乳癌危险的高度关联的结论。于是人们就提出了利血平致乳癌的“假设”，经进一步地验证这一假设，发现在对象选择方法上，对照组排除了心血管疾病的患者，而病例组并未排除，产生了偏倚的结果，导致利血平与乳腺癌的虚假联系。在之后的一项研究中避免了选择对象的不一致性，结果证明利血平与乳腺癌并无因果关联。从而纠正了这一“伪证”，“平反”了以上错误的结论。

（二）间接关联

当排除虚假关联后，不一定说明暴露因素与疾病肯定存在因果关联。当两类毫不相关的事件，如两种疾病都与某因素有关联时，这两种疾病会呈现明显的统计学关联，这种关联称为间接关联。为了避免与间接因果关联混淆，现在称其为继发关联（secondary association）。继发关联是由混杂偏倚引起的关联，即可疑的病因（暴露 A）与疾病 B 并不存在因果关联，但由于两者有共同的原因 C，因此观察到 A 和 B 与 C 均存在着关联。这也是在研究中应注意要避免的。比如有调查发现，有伤寒病史者的痢疾发生率明显比无伤寒病史者高，提示伤寒病史可能与以后发生痢疾相关。这种相关经假设检验有统计学意义，可排除抽样误差的可能；同时经仔细分析研究设计和所获资料，认为偏倚得到控制，可以排除虚假关联。但以现代医学理论解释，伤寒病史与之后的痢疾发生毫无相关，但两者均受到卫生状况及个人卫生习惯的影响。因此，认为两者的关联属于继发关联。再例如，高血清胆固醇是冠心病的危险因素，高血清胆固醇又可产生沉积于眼睑的黄色瘤，从而导致黄色瘤与冠心病的继发关联。这是一种纯粹由混杂偏倚产生的关联，即怀疑的病因黄色瘤与冠心病并不存在因果关系，而是由于两者均有共同的原因高血清胆固醇，黄色瘤和冠心病都与高血清胆固醇存在关联，从而导致黄色瘤与冠心病的继发关联。

当暴露因素与疾病 D 既存在直接关联，又存在间接（继发）关联时，暴露与疾病的直接因果关联的程度或方向将可能受到混杂干扰，即得到歪曲的关联估计值。例如，静脉吸毒（共用注射器）与性乱都是人类免疫缺陷病毒（HIV）感染的危险因素，吸毒者倾向于发生性乱行为，即吸毒同 HIV 感染既存在直接关联，又存在间接关联。在这种情况下，需要控制性乱的影响，避免性乱对吸毒和 HIV 感染的直接因果关联起混杂或歪曲作用。

当排除抽样误差、虚假关联和继发关联后，两事件间的关联才可能是因果关联，才能进行暴露因素与疾病的病因推导。

（三）因果关联（causal association）

统计学关联是判断因果关联的前提，但只有少数统计学关联属于因果关联。因果关联可以有直接关联和间接关联，随着研究的深入，直接关联和间接关联可以相互转化，原来认为是直接病因的可以被后来的研究证明是间接病因。我们可以用因果关系判断标准（病因学研究评价原则）来推断所研究的因素是否是疾病的病因。图 8-7 概括了进行病因推断前必须考虑的问题和步骤。

关于疾病的病因研究，一般言之，不能依靠临床观察的方法来研究病因。临床观察可以积累丰富的经验，它对病因问题可以提供某些疑点和线索，或经验性证据。例如，1959 年至 20 世纪 60 年代初期关于海豹短肢畸形与孕妇服用沙利度胺（thalidomide）的关系的研究，这种出生缺陷先在前联邦德国，后在许多欧洲国家流行，甚至在社会上引起了恐慌。尽管是由临床医生首先提出可疑之点，即孕妇服过缓解妊娠反应的沙利度胺，并怀疑可能同新生儿短肢畸形有关，

笔记

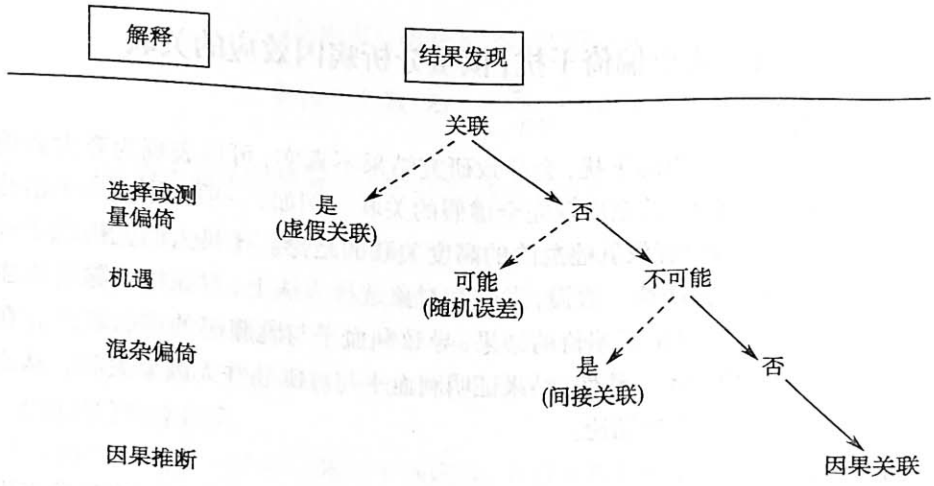


图 8-7　因果关联的判断进程

进而引起了公众关注。但这个重要的病因，却是通过有关国家疾病统计数字的累积，特别是经过分析性流行病学研究，才被最终认定。

人们往往有这样一种错误的看法，一提病因研究，就认为只是基础医学（即实验医学）的研究。其实不然，尽管实验医学的研究者可以利用现代科学技术，进行在有控制的条件下的某些动物实验、组织细胞实验以及分子水平的研究，以探索某因素与疾病的因果关系。但这些毕竟只能解释关于人类疾病过程中整个生态系统的某个个别阶段的变化和规律。人类是在复杂的社会和自然环境中生活的，这与实验室所控制的条件、环境完全不同。况且人与动物还有种属差异。因此，实验医学是重要的，但其研究结果往往不可轻易地引伸至人。再者，由于医德的原因，不能在人群中随意进行实验研究，因此，关于疾病病因问题的研究，很大程度上取决于以人为研究对象的流行病学研究或临床流行病学研究，且大多数是观察性研究。

第四节　疾病病因与危险因素研究的评价原则

疾病病因与危险因素研究结果是否能够确定病因，其研究的水平和价值多大，应对其进行严格的评价。表 8-5 所列是国际临床流行病学有关病因学研究的评价原则，其可以作为因果关系推断的参考标准。

表 8-5　病因学研究的评价原则

一、研究结果的真实性评价

1. 是否采用了论证强度高的研究设计方案
2. 除研究的暴露因素外，试验组与对照组其他方面是否一致
   - 包括 RCT、队列研究、病例对照研究
   - 其他已知的预后因素是否一致或经过了调整
3. 试验组和对照组有关因果效应的测量方法是否相同
   - 有回忆性偏倚、调查偏倚吗
   - 是否采用了盲法
4. 随访时间是否足够长？研究结果包含了所有随访病例吗
   - 随访不完全的原因
   - 失访病例与未失访病例的危险因素是否相似
5. 是否有因果效应的时间先后顺序
   - 暴露先于结果

笔记

续表

6. 有剂量 - 反应梯度关系吗
● 随着暴露剂量、持续时间增加，结果的危险性增加
7. 病因学研究的结果是否符合流行病学的规律
8. 病因致病的因果关系是否在不同的研究中反映出一致性
9. 病因致病效应和不良反应发生的生物学依据是否充分
二、研究结果的重要性评价
1. 暴露与结果联系的强度如何
● *RR*、*OR* 或 *EF*、*NNH* 等
2. 危险估计的精确性如何
● 可信区间（95% *CI*）
3. 研究的样本量合适吗
三、研究结果的适用性评价
1. 你的患者与文献中的研究对象是否存在较大的差异，以至于文献的研究结果不能回答患者的疑问
2. 你的患者发生疾病 / 不良反应的危险性有多大
3. 确定你的患者的喜好和希望解决的问题
4. 是否应终止接触暴露因素或更改治疗措施

一、病因学研究结果的真实性评价

（一）是否采用了论证强度高的研究设计方案

不同的病因学研究方法，其因果论证强度是不同的，描述性研究论证强度最弱，病例对照研究论证强度不太强，队列研究论证强度较强，随机对照的临床试验最强，因为它来自以人为研究对象的真正的人体实验。表 8-2 可作为评价不同的设计方案所获得的因果关系论证强度参考。

对于病因研究的证据是否来源于真正的人体实验，是指在研究中是否将人体置于暴露环境中，通过比较不同组别的结局来判断暴露因素的致病效应。临床上多使用随机对照试验，具有较强的因果论证强度。但如果在健康人体中进行施加危险因素的病因学实验，显然是不道德的，并且不具有可行性。但可以变通的采用去除可能的致病危险因素的方法来进行干预性研究设计。这种所谓的终止效应可以来自实验流行病学研究、自然实验或自发性改变的观察性资料。是否存在终止效应也是一个很关键的标准，也即如果对某些可能的致病危险因素或病因采取针对性的措施之后，能使被研究的疾病发生率下降或流行随之减少，那么对病因或危险因素的肯定也具有非常重要的意义。

（二）除研究的暴露因素外，试验组与对照组其他暴露是否一致？是否存在混杂因素影响

混杂因素可能是已知的，也可能存在未知者，其对病因研究影响颇大，因此，在拟探讨的致病因素之外，应该注意在试验组和对照组中是否存在混杂以及混杂的程度，是否采用了适当的控制或消除办法。

（三）组间对因果效应的测量方法是否一致？是否采用盲法

在病因学研究中，对两组研究对象（暴露组与非暴露组）应采用同样的观察或调查方法、对暴露（不良反应）结果应使用相同的测量手段和指标。且应采用盲法来评价暴露结果，这样才有利于防止测量性偏倚干扰，使各项因果测试的方法，指标一致有利于对比，避免上述偏倚造成的影响，增加结果的真实性。

（四）随访观察时间是否足够长？研究结果是否包括了所有纳入随访研究的病例

任何致病因子引起人体发病都有一个致病的时间效应关系，急性及自然病程短的疾病则致病效应期短，如急性传染病。对于慢性非传染性疾病则其致病效应期较长，因此，研究慢性非传染性疾病发病危险因素的致病效应时，往往需要足够时间才能观察到结果的发生，观察期过短会

笔记

获得假阴性结果。另外，在随访期间失访过多会影响研究结论的真实性（一般失访率不应 >20%），因为中途退出的研究对象可能在某些特征上与仍然留在研究中的研究对象存在差别。

（五）因果效应的时间顺序是否确切、合理

致病因素引起发病，必然是因在前，果在后，时序性是构成因果关系的基础。在评价某一病因学或不良反应研究时，如果能明确危险因素的出现早于疾病或不良反应的发生，则研究结果的真实性就高。对于实验性研究和队列研究可以确定因果效应的时间顺序。而横断面的、回顾性的以及叙述性研究则不能确定因果效应的时序性。因此在确定病因致病的因果顺序时应选用灵敏度高、特异性好的测量指标，制定明确的诊断标准，证明接触病因前被研究对象未患有所研究的疾病。

（六）是否存在剂量 - 效应的致病关系

剂量 - 效应关系是指暴露因素的剂量、程度与疾病发生的进展与程度存在显著的相关关系。常见于化学性和物理性有害因素的损害作用，其致病效应的程度与接触的剂量存在相关，接触的剂量越大，以及接触的时间长，累积剂量大，致病效应越明显，损害越大。如果能够证明随着与可疑致病因素的接触增加（剂量增大、时间延长）导致更多更重的相关疾病的发生，则就能够肯定这一效应关系。

（七）病因学研究结果是否符合流行病学的规律

流行病学在致病因素、机体和环境等方面宏观地进行探讨，为病因学研究常能提供重要的发现和有力的证据。研究中的暴露的分布，如与疾病的地理分布、时间分布及人群间分布符合或基本符合，则意义更大。

（八）不同的研究，结论是否一致

对于某一疾病的病因，在不同地区、不同人群中采用不同的或者是相似的研究方法，并采用相关的评价指标进行的病因研究，所得到结论如果均一致，那么该病因致病的结论就较为可靠，循证医学开展的系统评价所提供的信息是评价结论是否一致的重要依据。它是病因推断中一个重要的指标，符合 Mill 准则中的求同法则。

例如，吸烟与肺癌关系的病因学研究，世界上至少已有数十项病例对照研究和 7 次以上较大规模的队列研究的结果均表明，肺癌与吸烟有很强的关联。多次研究的可重复性使因果关联的可能性增加，而少数或个别研究的不同甚或相反的结果并不能简单否定，需要仔细探究结果差异的缘由。由于临床病因学研究是一项较为复杂的过程，在研究设计及实施过程中不同的研究者会存在差异，因此所得到的结论并非都能反映真正的因果关联，因此在病因的一致性分析中要持谨慎的态度，权衡研究设计、实施过程及其结果的科学性。

（九）病因学效应的生物学依据是否充分

如果病因学研究揭示的因果关系有生物学的可解释性，则可增加因果联系的论证强度。随着当今生命科学研究的飞速发展，把临床流行病学对致病因素的宏观研究结果与基础医学的分子生物学、细胞生物学、分子病理学、组织学水平、遗传学和免疫学等微观研究结果相结合，必将促进病因学研究的飞速、深入发展，将会对疾病有更加清晰的认识。

二、病因学研究结果的重要性评价

通过上述真实性的分析与评价，假设一个病因学研究的结果有着良好的真实性，那么我们应该进一步地评价这个研究结果有否具有重要的临床或公共卫生学意义和价值，否则就没有必要再评价了。在评价重要意义方面则应有量化的相应指标。

（一）关联的强弱程度

在病因学研究的方法中有一些反映暴露与疾病的因果关联程度的量化指标。随机对照试验以及队列研究中，常用相对危险度（*RR*）、归因危险度（*AR*）、*RRI*、*NNT* 等来评价因果关联强

笔记

度，病例对照研究则多用优势比（*OR*）进行评价。

在判断 *RR* 和 *OR* 的意义时，还要进行敏感性分析（sensitivity analysis），这有助于对潜在的混杂因素进行"调整"或"修正"。如果调整后的 *RR* 或 *OR* 较调整前的 *RR* 或 *OR* 明显变小，则应该怀疑原来的结果。相反，如果调整后的 *RR* 或 *OR* 与调整前相比保持不变或比调整前明显增大，则可以更确信该因果关联的真实性。

（二）相对危险度（*RR*）或优势比（*OR*）的 95% 可信区间

除评价因果关系的点估计强度外，还需评价其精确度，方法是计算 *RR* 或 *OR* 等指标的 95% 可信区间，如果 95% *CI* 的范围较狭小、则其精确度就高，下限和上限值不包括 1.0，表明有统计学意义。如果研究发现暴露因素与疾病暴露因素与疾病 / 不良反应存在相关性，则可信区间下限表示相关强度的最低值，如果研究未发现两者的关系（没有统计学意义），则可信区间的上限表示相关强度可能达到的最高水平。

（三）害 - 需暴露人数

害 - 需暴露人数（number needed to harm，*NNH*）是指多少人暴露于危险因素 / 干预措施，就可能比非暴露组多发生 1 例疾病 / 不良反应。*NNH* 是一种更直观、更易被临床医生理解的表示因果相关强度的指标。*NNH* 为 *AR* 的倒数，即 $NNH=1/AR$。在疗效评价不良反应研究中，则可称之为害 - 需治人数。在疗效评价研究中，其相对应的指标为 *NNT*（number needed to treat，益 - 需治人数）。

三、病因学研究结果的适用性评价

在评价了病因和危险因素研究结果的真实性和重要性后，如果这些研究证据具有良好的真实性，而且又具有临床或公共卫生学的重要意义，那么应联系患者的实际情况来探讨病因和危险因素，以助于解决患者的实际问题，所以我们还需考虑该研究证据的结果是否能够用于我们自己的患者或公共卫生实践。

（一）你的患者与文献中的研究对象是否存在较大的差异，以至于文献的研究结果不能回答患者的疑问

当考虑研究证据的适用性时，首先要比较你的患者与研究证据中的研究对象在年龄、性别、经济收入、文化背景等社会学特征及疾病严重程度、病程、用药剂量、依从性、对治疗的反应性等临床特征等方面的差异。如果存在影响结局的重要差异，可能研究证据就不能用于解决你的患者的问题，或具体实践。

（二）你的患者发生疾病 / 不良反应的危险性如何

针对单个病人的问题，需要根据研究证据提供的信息，估计你的患者发生疾病 / 不良反应的危险性。一种方法是在文献中寻找与你的患者各方面特征比较一致的亚组，参照该亚组的 *NNH*，但在这种情况下，一定要慎重，因为亚组的样本量常常较少，受机遇的影响较大。另一种方法是，计算患者发生疾病 / 不良反应的可能性与文献报告可能性的比值 *F*（decimal fraction），然后用文献结果中的 *NNH* 除以 *F* 值，获得你的患者发生疾病 / 不良反应的危险性大小。

（三）确定你的患者的喜好和希望解决的问题

医务人员在进行医疗决策时，除了参考研究证据提供的信息，坚持利大于弊的原则同时还应邀请患者参与医疗决策。病人参与医疗决策是为了尊重患者的权利，不同的病人因其对自身疾病的关心程度，对医生所给予的诊治措施的期望值及对不良反应的耐受性等的不同，最终的选择会有差别。我们可以调查患者的喜好，提供一系列代表危险性和良好结局的价值尺度，了解患者在哪一点时会改变他们的决定，即接受或拒绝治疗措施（暴露）。

（四）是否应终止接触暴露因素或更改治疗措施？

如果研究证据显示暴露因素与疾病 / 不良反应间具有较强的因果联系，即接触危险因素的

笔记

危险明确而且巨大，决策就很明确，即立即脱离接触危险因素或终止治疗措施。对于不良反应因果关系研究来说，即使证据显示暴露因素（治疗措施）与疾病/不良反应间因果相关性不是太强，但有其他药物可选也容易作出决策。在临床实践中，同一种疾病常常有多种治疗措施可供选择。因此，医生应告知患者和与患者讨论每一种措施的利弊，最后决定采用哪一种治疗措施。

以上所列病因学研究和评价的依据，实际应用中可以根据研究要求综合参考，灵活运用。可以根据不同的研究设计，有所侧重地运用其中的条款。另外，上述原则也可作为阅读和分析病因学的文献时，鉴别真伪的参考尺度。

第五节 循证实践及案例分析

一、疾病案例

患者女性，68岁。以“跌倒1小时”为主诉入院。入院当晚在夜间起床去卫生间过程中跌倒。入院后X线检查，发现“左侧股骨嵌插性骨折”。在骨科进行了内固定手术。经询问病史得知，患者有睡眠障碍10年，长期睡前服用苯二氮䓬类药物（艾司唑仑）。因此，主治医生希望了解患者夜间跌倒发生骨折是否与长期服用苯二氮䓬类药物有关。

二、提出问题

医生在给患者采用任何干预措施前需要明确，该干预措施可能引起的不良反应，其造成的伤害是否小于治疗带来的益处，尤其在面对不同个体时，患者的基础状况、特异性体质和合并用药情况均需要评估，以考虑是否会增加治疗的不良反应。对于本病例，患者有睡眠障碍，且一直在服用苯二氮䓬类药物，主治医师希望了解长期服用苯二氮䓬类药物是否会增加患者骨折的风险，这显然是一个药物不良反应的问题，是了解长期服用苯二氮䓬类药物与骨折之间是否存在因果关联的问题，从循证实践的角度，要明确这个问题就必须进行全面的文献搜索，以了解相关的证据。基于这个初始临床问题，为了能够准确检索相关的证据，需要进行相应的问题转换，构建相应的临床问题。

目前，通用的构建临床问题的方法是遵循PICO原则，即构建问题的四个要素：研究对象（patient）、干预措施（intervention）、与干预措施相比较的措施（comparison intervention）和产生的结局（outcome）。本例，依照该原则构建过程见表8-6。

表8-6 遵循PICO原则的构建过程

P：患者	睡眠障碍患者
I：干预措施	长期服用苯二氮䓬类药物
C：比较措施	不用药
O：结局	骨折

经过转换后的问题是：长期服用苯二氮䓬类药物的患者与未服用该药的患者相比，是否会增加发生骨折的风险？

三、证据检索与评价

浩如烟海的医学信息中，要快速地获得所需要的资料，需要掌握文献检索的方法和技巧，包括检索词以及文献数据库的选择、检索策略的制定等。由于不同文献资料的质量不同，研究的设计和实施、数据的整理与分析、结果的解释和论文报告等方面存在差异，研究的真实性和可靠性以及适用性也不同。在循证医学实践中，建议首先检索二次数据库，即经他人评估和筛

笔记

选过的循证医学资源，如 Cochrane 图书馆、Clinical Evidence 等，如未检索出所需要的信息，再进一步检索未经筛选的数据库，如 PubMed、Springerlink、CBM、中国期刊全文数据库等。

针对上述临床问题，具体文献检索策略的制定请读者参照本教材相关内容，本章仅以"Benzodiazepine"和"fracture"（中文数据库以"苯二氮䓬类"和"骨折"）为检索词进行检索，检索日期为2012年7月18日。检索结果见表8-7。

表8-7 检索结果

数据库	检索结果	与本例临床问题相关
PubMed	219	39
CBM	1	0
中国学术期刊网	7	0
维普资讯	1	0
万方数据库	2	0
OvidSP	11 755	36
Springerlink	2835	16
Wiley	1549	12
Clinical evidence	6	6
ACP Journal Club	1	1
The Cochrane Library	35	7

循证实践在回答临床问题时，应首先选择针对该问题的系统评价（systematic review），因为系统评价客观总结了所有相关的文献资料，其作为证据的论证强度最高。本例经去重后，共在国内外的11个数据库中检索到相关文献70篇，经过进一步评价筛选，本章分别选取队列研究和系统评价各1篇，分别是 Anita K. Wagner 等的队列研究 *Benzodiazepine Use and Hip Fractures in the Elderly——Who Is at Greatest Risk?*（Arch Intern Med，2004，164：1567-1572）及 T.P. Khong 等的系统评价 *Potential Impact of Benzodiazepine Use on the Rate of Hip Fractures in Five Large European Countries and the United States*（Calcif Tissue Int，2012，91：24-31），参照表8-2分别对其真实性、重要性和适用性进行评价，以指导循证实践，回答提出的临床问题。

四、实践决策

（一）队列研究：*Benzodiazepine Use and Hip Fractures in the Elderly——Who Is at Greatest Risk?*（Arch Intern Med，2004，164：1567-1572）

1. 真实性评价

（1）是否采用了论证强度高的研究设计方法：依据表8-2，在病因研究设计类型上，以随机对照试验（RCT）的论证强度最高，其次是队列研究。通过检索得知，目前关于此临床问题的RCT研究较少，大多为队列研究和巢式病例对照研究，而本研究即属前瞻性的队列研究，论证强度较高。

（2）除研究的暴露因素外，试验组与对照组其他方面是否一致：本研究为队列研究，尽管不能像随机对照试验一样严格控制组间的可比性，但暴露组与非暴露组均来自同一队列，因此组间的可比性应该是有保证的。

（3）试验组和对照组有关因果效应的测量方法是否相同（结果测量是否客观或采用盲法）：本研究数据来源于 New Jersey Medicaid health care claims data，根据以往和当前的药物使用和诊断信息，将研究对象分配到苯二氮䓬类药物暴露组和非暴露组。骨折的诊断则根据国际疾病分类代码（第9版）820.xx进行。因此，本研究暴露组和对照组的测量方法应当是客观且一致的。研究中暴露因素和结局的测量均采用标准化的方法，因此不存在盲法的问题。

（4）随访时间是否足够长？研究结果包含了所有随访病例吗：文中表明，研究者对研究对象

笔记

进行了为期42个月的随访。研究共对125 203名研究对象进行了随访，报告的研究结果包含了所有的随访病例。

（5）是否有因果效应的时间先后顺序：本研究属前瞻性队列研究，特点就是在特定人群中按照目前或过去是否暴露于待研究的危险因素，分为暴露与非暴露组，随访观察一段时间后，进而比较待研究疾病的发病率。因此，因果时相关系是明确的。

（6）是否存在剂量 - 反应梯度关系：文中虽提到了在收集资料的过程中采用地西泮当量（mg）统一评估药物使用剂量，但在结果分析过程中并未涉及剂量 - 反应梯度关系。

（7）病因学研究的结果是否符合流行病学的规律：关于暴露因素（苯二氮䓬类药物使用）与疾病 / 不良反应（骨折）是否存在消长关系，本文中未提及。

（8）病因致病的因果关系是否在不同的研究中反映出一致性：本研究结果显示，服用苯二氮䓬类药物会使发生骨折的风险增加0.24倍（*RR*＝1.24，95% *CI*：1.06~1.44），与诸多研究如Lichtenstein（1994），Wang（2001），Cummings（1995），Chang（2008）等一致。但也有部分研究结果未得出这一结论，如Ensrud（2003），Ray（1989），Sgadari（2000）等研究。因此，需进一步对此类研究进行系统评价，通过增大样本含量，减少随机误差所致的差异以提高统计效率；探讨多个研究结果间的异质性，实现不一致的研究结果的定量综合估计。

（9）病因致病效应和不良反应发生的生物学依据是否充分：此种因果关系的生物学依据至今尚不明了，文中也未提及。

2. 重要性评价

（1）关联的强度（*OR*或*RR*）：与未服用苯二氮䓬类药物（BZD）组相比，服用BZD组的调整*IRR*值为1.24，长半衰期BZD组为1.13，短半衰期高效BZD组为1.27，短半衰期低效BZD组为1.22，服用超过一种BZD组为1.53，服用少于15天BZD组为2.05，服用16~30天BZD组为1.88，持续服用BZD组为1.18。

（2）相对危险度（*RR*）或优势比（*OR*）的95%可信区间：与上述*IRR*相对应的95%可信区间分别为（1.06~1.44），（0.82~1.55），（1.01~1.59），（0.89~1.67），（0.92~2.53），（1.28~3.28），（1.15~3.07），（1.03~1.35）。除长半衰期BZD组，短半衰期低效BZD组，服用超过一种BZD组外，均有统计学意义。

（3）害 - 需暴露人数（*NNH*）：本例中随访资料以人年为单位，因此可以发病密度来计算*NNH*，以服用BZD组为例，$NNH = 1/(347/19\,956.0 - 1965/174\,115.0) = 163.88$，即每暴露163.88人年会发生一例骨折。

3. 适用性评价

（1）你的患者与文献中的研究对象是否存在较大的差异：由文中的一般人口学描述中可知，本例中研究对象为65岁以上老年人，因此，研究结果应该适用于该患者。

（2）患者发生疾病 / 不良反应的危险性如何：根据患者的情况与文中对比，该患者应属于持续服用BZD亚组，该组$NNH = 1/(305/18\,290.8 - 1965/174\,115.0) = 185.56$。该亚组人数为24 042人，随访人时数为18 290.8人年，样本量较大，应可避免受机遇的影响。

（3）是否应终止接触暴露因素或更改治疗措施：研究证据显示该暴露因素与不良反应间具有一定的因果相关性，但关联强度较弱。但患者睡眠障碍的治疗也不应忽视。应与专科医生会诊后，结合病人的意愿进行决策。

（二）系统评价：*Potential Impact of Benzodiazepine Use on the Rate of Hip Fractures in Five Large European Countries and the United States*（Calcif Tissue Int，2012，91：24-31）

1. 研究结果的真实性

（1）是否是根据随机对照试验进行的系统评价：是否纳入高质量的原始文献，是评价系统评价作为证据的论证强度高低的重要标准。随机对照试验能够较好地控制各种偏倚对结局的影

笔记

响，是评价干预措施疗效的最佳设计方案，因此依据随机对照试验产生的系统评价，其研究证据的论证强度被认为最高。但该方法多用治疗措施的有效性评价，而限于伦理学的要求，不良反应研究则较少应用。通过检索也得知，目前关于此临床问题的 RCT 研究较少，大多为前瞻性的队列研究和巢式病例对照研究，本系统评价研究为包含欧洲五个国家（法国、德国、意大利、西班牙和英国）和美国在内的多个研究的系统评价，是对多个队列研究和巢式病例对照研究的定量综合。

（2）在系统评价的"方法学"部分，是否描述了检索和纳入所有相关研究的方法？评价单个研究证据的方法：文中清晰的说明研究者以 hip fractures 和 benzodiazepines 相关词条为检索词，于 2010 年 9 月和 10 月在 PubMed、the Cochrane Library 和 Embase 等数据库进行了系统检索，而未公开发表的数据未纳入研究。

文中虽未提及评价单个研究证据的具体方法，但明确给出了 5 条纳入标准以保证单个研究的质量。

（3）不同研究的结果是否一致：所纳入的原始文献中，关于长期服用苯二氮䓬类药物是否会增加发生骨折的风险，结论莫衷一是。但系统评价 /meta 分析的主要作用即是对具有相同研究目的的多个研究结果进行综合定量分析的一种方法，其优点是通过增大研究样本含量，减少随机误差所致的差异以提高统计效率；探讨多个研究结果间的异质性，实现不一致的研究结果的定量综合估计。本系统评价共纳入了 11 项研究的 14 组数据，其中 8 组数据为阳性结果，即服用苯二氮䓬类药物会增加骨折发生的风险，而另外 6 组数据未能得出此结论。另外，本文还采用异质性检验的方法来判断不同研究的结果的异质性。结果显示，I^2 值为 42%~66%，表明存在一定的异质性。

（4）统计分析用的数据资料是单个患者的资料还是单个研究的综合资料：系统评价的资料来源可以是原始研究中群体资料，也可以是原始研究中的个体资料，两种类型各有优缺点，应根据研究要求和已有条件进行选择。根据文中的内容可知，本系统评价统计分析中所使用的数据资料为单个研究的综合资料。

2. 研究结果的重要性

（1）发生不良反应的强度大小如何（效应测量指标的大小）：本研究中通过对多个研究进行 meta 分析后，直接给出了合并后的 *RR* 值为 1.40，表明服用苯二氮䓬类药物发生骨折的风险是不服用苯二氮䓬类药物的 1.4 倍。而原始文献的提供的 *RR* 值为 1.09~2.05 不等。长效苯二氮䓬类药物的 *RR* 值为 1.32，短效的 *RR* 值为 1.23。

（2）发生不良反应的精确性如何（可信区间的大小）：与单个研究结果一样，系统评价结果的可信区间可以提供关于研究结果精确性的信息，若可信区间较宽，则表示精确性较低，反之则表示研究结果的精确性较高。本文中也直接给出了 95% 可信区间，为（1.24~1.58）。长效和短效药的 95% *CI* 分别为（1.10~1.58）和（1.09~1.39）。

3. 研究结果的适用性

（1）你的病人是否与系统综述中的研究对象差异较大，导致结果不可用：由纳入的文献可知，多数文献中的研究对象为 65 岁以上老年人，因此，研究结果应该比较适用于该患者。

（2）规避不良反应的措施在你的医院是否可行：苯二氮䓬类药物是临床治疗睡眠障碍的常规用药之一，但不是唯一治疗用药，因此可以根据医生的经验和病人的取向选用其他替代药物。

（3）你的病人从治疗中获得的利弊如何：选用其他替代药物，在控制患者睡眠障碍的同时，又能够有效减少骨折等不良反应的发生，患者无疑将从中获益。

（4）对于治疗的疗效和不良反应，你的患者的价值观和选择如何：临床医生应该根据自己掌握的证据，耐心细致的向患者阐述更换治疗方法的利与弊，最后结合患者的价值观和选择，进行临床决策。

笔记

五、小　结

综上所述，该患者骨折的发生或与长期服用苯二氮䓬类药物有关，应建议改用其他治疗睡眠障碍药物，并做好防范措施，防止跌倒等。

（孙业桓）

笔记

第九章　诊断试验研究及评价与实践

诊断试验涉及临床症状、特殊体征、各种实验室检查以及特殊检查，是临床诊断与鉴别诊断的重要工具。开展诊断试验的目的是寻找简便、安全、经济并为患者易接受的方法用以替代复杂、高风险、高成本或者临床操作困难的检查措施。最佳研究设计为采用与金标准方法同期盲法比较的诊断试验。诊断试验评价指标包括敏感度、特异度、预测值、似然比及受试者工作曲线。应用诊断试验结果进行循证医学实践时，应注意评估诊断试验研究的科学性、重要性以及适用性。

第一节　研究和评价诊断性试验的意义

诊断试验（diagnostic test）可为疾病正确诊断以及鉴别诊断提供重要证据。诊断试验的实施，要涉及以下内容：①病史和体检所获得的临床资料；②各种实验室检查，如生化、血液学、细菌性、免疫学、病理学检查等；③各种影像学检查，如X线造影、B超、CT、磁共振成像（MRI）及放射性核素等；④其他特殊器械检查，如心电图、内镜等；⑤各种临床公认的诊断标准：如各种自身免疫性疾病的联合诊断标准等。

诊断试验主要用于诊断或者排除某种疾病，其他用途还包括：①判断疾病的严重程度；②估计疾病的临床过程、治疗效果及其预后；③筛检无症状的病人；④监测药物不良反应。

尽管目前常见疾病的诊断标准与诊断措施已先后确立，但由于新疾病、新诊断方法的不断涌现，仍不能满足临床对诊断研究证据日益增长的需求，特别是在以下几个方面：①新的疾病缺乏诊断"金标准"，需要发展诊断试验并对其进行评价；②新的诊断方法，如各种新的影像检查、各种分子生物学检测方法是否优于现有的方法，是否适合在临床开展，其安全性和经济性等都需要评估；③多种诊断试验如何合理安排、次序如何确定也需科学地决策。因此，诊断试验研究与评价的意义及其临床重要性是显而易见的。

第二节　诊断试验研究的基本方法

评价一项新诊断方法的诊断价值，最佳研究设计是采用与金标准（gold standard）方法同期盲法比较的诊断试验。第一，确立疾病的标准诊断方法（最好是金标准），第二，选择研究对象，并根据标准诊断将这些对象划分为"有病组 - 病例组"与"无病组 - 对照组"，第三，再用新诊断方法同步测试这些研究对象，将获得的诊断结果与标准诊断比较，进而绘制四格表，计算敏感度、特异度、预测值、似然比等指标来综合评价该试验的诊断效率和价值。为减少偏倚，一般采用盲法评价。

一、确定诊断试验评价的金标准诊断

标准诊断，又叫做"金标准"诊断，是指当前临床医学界公认的、诊断某疾病的最可靠方法，常用的金标准诊断方法有病理学诊断（组织活检和尸体解剖）、外科手术发现、特殊影像学诊断（如冠状动脉造影诊断冠心病），也可采用公认的临床诊断标准（如系统性红斑狼疮的ARA诊断标准等），通过长期临床随访所获得的肯定诊断，有时也可用作金标准诊断。

笔记

开展诊断试验时，金标准诊断的选择应结合临床具体情况而定。例如肿瘤诊断的金标准是病理诊断，冠心病诊断最好选用冠状动脉造影，胆石症诊断则一般以手术发现为准。若金标准选择不当，就会造成“病例组”和“对照组”的划分错误，从而影响诊断试验的正确评价。同时，临床诊断试验还要遵循临床研究及伦理学的一些基本原则，如不损害患者原则、伦理委员会通过原则、患者知情同意原则等，变相增加了金标准操作的难度，有时要根据具体临床疾病，选择临床公认且得到广泛应用的诊断方法作为金标准。如诊断冠心病的金标准是冠脉血管造影，但造影有一定风险、操作存在一定难度，尤其是对照组的研究对象，此时可采用较为成熟的冠脉CT成像技术（CTA）作为金标准来使用。因此，临床研究中的金标准是相对的，但应注意：这些相对的金标准诊断，可能会带来一些偏差，需要一定的方法校正结果。

二、诊断试验研究对象的选择和样本量的确定

诊断试验的研究对象应具有代表性，其中病例组应包括各种临床类型和处于不同病程阶段的病人：如轻、中、重型，早、中、晚期，典型的和不典型的，有和无并发症者，已治疗与未经治者等。对照组应选自确未罹患该病的其他病例，尤其应包括易与该病相混淆的病例，设置这样的对照才具有鉴别诊断的价值。在试验研究初期，有时也可选正常人作为对照组。研究对象应同期进入研究，可以是连续样本或者是按比例抽取的样本，但不能由研究者随意选择，否则就会出现选择偏倚，影响试验的真实性。

诊断试验研究的样本量与下列因素有关：①对试验敏感度的要求，即假阴性率要控制在什么水平，敏感度高的试验一般用于疾病的筛选；②对试验特异度的要求，即假阳性率要控制在什么水平，特异度高的试验一般用于肯定诊断；③允许误差，一般取总体率 $100(1-\alpha)\%$ 可信区间宽度的一半。

计算公式：$n=[u_\alpha^2 p(1-p)]/\delta^2$。

式中 n 为所需样本大小，u_α 为正态分布中累积概率为 $\alpha/2$ 时的 u 值，一般取 1.96，δ 为允许误差，一般定在 0.05~0.10 之间；p 指敏感度或特异度，可采用敏感度的估计值来计算病例组所需样本量，用特异度的估计值来计算对照组的样本量。

三、绘制四格表

诊断试验的设计，首先应绘制四格表（表 9-1）。纵列为根据金标准诊断划分的“病例组”及“对照组”结果，而待评价的诊断试验结果则横向排列。当同步检测这两组对象后，将阳性结果列为诊断试验的第一行，阴性结果列为第二行，由此构成了一个四格表。

表 9-1 诊断试验评价四格表

诊断试验	金标准诊断方法评估结果		合计
	病例组	对照组	
+	a 真阳性	b 假阳性	$a+b$
−	c 假阴性	d 真阴性	$c+d$
合计	$a+c$	$b+d$	N

a：真阳性，为病例组内试验阳性的例数；b：假阳性，为对照组内试验阳性的例数；c：假阴性，为病例组内试验阴性的例数；d：真阴性，为对照组内试验阴性例数；N：总人数。

各项评价指标的计算公式：

敏感度 $=a/(a+c)$；特异度 $=d/(b+d)$；准确度 $=(a+d)/N$；患病率 $=(a+c)/N$；

笔记

阳性预测值：$a/(a+b)$；阴性预测值：$d/(c+d)$；

阳性似然比：$[a/(a+c)]/[b/(b+d)]$；阴性似然比：$[c/(a+c)]/[d/(b+d)]$。

四、盲法比较诊断试验结果

在合理选择金标准的同时，应使用盲法独立评价诊断试验和金标准的诊断结果。即评价者事先无法获知研究对象的具体分组情况，是"有病"组还是"无病"组，同时评价同一病人的诊断试验与金标准结果也是独立进行的。这里有三层意义：一是准确性验证，无论待评的诊断试验或者金标准诊断，都需要确定其结果判定的准确性或者稳定性，至少应有2个独立专家进行评价，结果不一致时，再通过第三位专家或者讨论确定。二是独立性，即诊断试验和金标准诊断是独立进行的、互不干扰，顺序先后或者时间间隔也不会对结果判断产生影响。三是盲法比较，即对诊断试验和金标准诊断结果的判断也是相互独立的，在某些情况下，若获知了金标准结果，会影响甚至左右对诊断试验结果的判断。例如，评价者先在患者CT片上看到了肺部结节，在读X线胸片时，会先入为主，读片会更加仔细，甚至发现以前读片时并没有发现的小阴影，导致测量偏倚。

第三节　诊断试验的评价指标及其临床应用

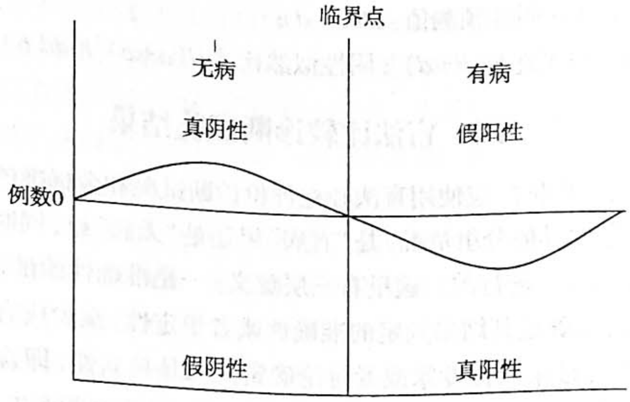
评价诊断试验的临床价值应从三方面考虑：真实性、精确性和适用性，其中以诊断试验本身的真实性大小最为重要，下列一些指标可用来评价诊断试验的真实性。

一、敏感度与特异度

敏感度（sensitivity），又称灵敏度，是指由标准诊断法确诊有该病的病例组中经诊断试验查出阳性人数的比例$[a/(a+c)]$。而病例组中诊断试验结果为阴性的即为假阴性患者，占病例组的比率就是假阴性率又称漏诊率，敏感度和假阴性率是互补的，即敏感度＝1－假阴性率。特异度（specificity）是指由标准诊断法确诊无病的对照组中经诊断试验检出阴性结果人数的比例$[d/(b+d)]$。而对照组中试验结果为阳性者即为假阳性，假阳性例数占对照组的比率就是假阳性率又称误诊率，特异度和假阳性率也是互补的，即特异度＝1－假阳性率。

敏感度和特异度都为100%的诊断试验最为理想，即没有假阳性和假阴性出现，有病组和无病组中结果数据分布曲线没有重叠（图9-1a），这在临床实践中往往难以实现，多数诊断试验结果的分布曲线是两组有部分重叠（图9-1b）。因此，诊断试验结果若为连续性变量时（如空腹血糖和血压等），区分正常（阴性）、异常（阳性）的临界点（cut-off point）划分很关键，将直接影响敏感度和特异度，如选择A点（图9-1b），该试验的特异度为98%，但敏感度仅为60%，造成许多病人漏诊；若要提高敏感度，可将临界点移向C点，敏感度可达95%，但特异度降低为60%，又造成许多误诊病例。因此敏感度和特异度一般呈反比关系。临床实践中可通过以下途径选择临界点：第一，可利用受试者工作特性曲线（receiver operator characteristic curve，ROC曲线）寻找合适的临界点，以避免过多的假阳性和假阴性；第二，根据临床需要，通过权衡假阳性和假阴性造成的后果，选择临界点，以达到临床需要的高敏感度或者高特异度。高敏感度试验适用于：①疾病严重但又是可治疗的，疾病的早期诊断将有益于病人，而疾病漏诊可能造成严重后果者，例如结核病，霍奇金病等；②有几个诊断假设，为了排除某病的诊断；③用于筛检无症状病人而该病的发病率又比较低，因此当试验结果呈阴性时，高敏感度试验的临床价值最大。高特异度试验适用于：①凡假阳性结果会导致病人精神和肉体上严重危害时，例如诊断病人患癌，而准备实施化疗；②要肯定诊断时，高特异度试验阳性结果的临床价值最大。

笔记



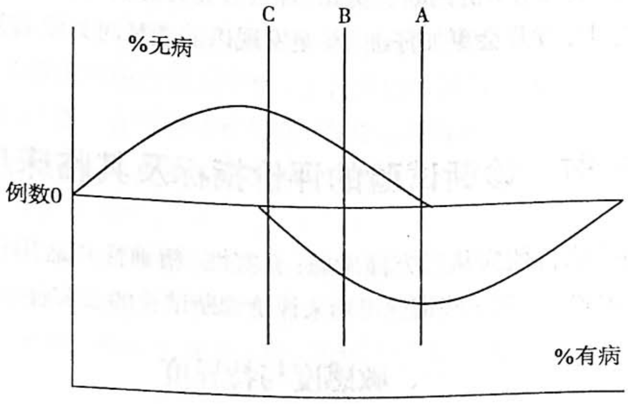


图 9-1　敏感度和特异度的关系

二、准　确　度

准确度(accuracy)又称总符合率、粗一致性(crude agreement rate),表示观察值与标准值或真实值符合的程度。作为诊断试验的评价指标,它是真阳性与真阴性之和占受检总人数的百分率,反映正确区分患者与非患者的能力,准确度高,则真实性好。

三、预测值与患病率

在临床实践中临床医师和病人都会关心诊断试验的诊断价值,如是阳性结果,患某病的可能性有多少,阴性结果时确未患病的可能性又有多少,这就涉及预测值(predictive value, PV)问题。阳性预测值(positive PV)是指试验阳性结果中真患该病的比例[$a/(a+b)$],阴性预测值(negative PV)是指试验阴性结果中确未患病的比例[$d/(c+d)$]。一般说来,越灵敏的试验,其阴性预测值越高;反之特异度越高的试验,其阳性预测值越高。预测值的高低并不完全取决于试验本身,还与患病率[$(a+c)/(a+b+c+d)$]有关。不同临床情况下,患病率有时相差甚大。例如,AFP 诊断肝癌试验若在某三级医院消化科应用,该诊断试验的敏感度和特异度分别为 80% 和 90%。如果在研究人群为肝硬化患者中进行,肝癌患病率大致为 50%,见表 9-2,AFP 诊断肝癌阳性预测值为 80/(80+10)=88.9%;如果在慢性乙型肝炎人群应用此诊断试验,人群中的肝癌患病率小于 10%(假设为 9.1%),表 9-3 显示,阳性预测值为 44.4%;如果在携带乙肝病毒人群中应用该诊断试验,此时的人群患病率不足 1%(假设为 0.99%),表 9-4 显示,阳性预测值为

笔记

7.4%；如果将AFP作为肝癌高发区普通人群（肝癌患病率100/10万人，0.1%）的筛查，表9-5显示，阳性预测值降至0.8%。由此可见，患病率对预测值有着非常重要的影响。因此，在应用或者引用文献报道的诊断试验时，应考虑其患病率是否与本医院或者当地情况相似，在三级医院阳性预测值很高的试验，可能在一级医院就很低。

表9-2 AFP诊断肝癌的临床研究评价（人群患病率50%）

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	10(10%)
AFP(−)	20(20%)	90(90%)
合计	100	100

人群患病率＝100/(100＋100)＝50%，敏感度＝80%，特异度＝90%，则：阳性预测值＝80/(80＋10)＝88.9%，阴性预测值＝90/(20＋90)＝81.8%。

表9-3 AFP诊断肝癌的临床研究评价（人群患病率9.1%）

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	100(10%)
AFP(−)	20(20%)	900(90%)
合计	100	1000

人群患病率＝100/(100＋1000)＝9.1%；敏感度＝80%，特异度＝90%，则：阳性预测值＝80/(80＋100)＝44.4%，阴性预测值＝900/(20＋900)＝97.8%。

表9-4 AFP诊断肝癌的临床研究评价（人群患病率0.99%）

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	1000(10%)
AFP(−)	20(20%)	9000(90%)
合计	100	10 000

人群患病率＝100/(100＋10 000)＝0.99%；敏感度＝80%，特异度＝90%；则：阳性预测值＝80/(80＋1000)＝7.4%，阴性预测值＝9000/(20＋9000)＝99.8%。

表9-5 AFP诊断肝癌的临床研究评价（人群患病率0.1%）

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	10 000(10%)
AFP(−)	20(20%)	90 000(90%)
合计	100	100 000

人群患病率＝100/(100＋10 000)＝0.1%；敏感度＝80%，特异度＝90%；则：阳性预测值＝80/(80＋10 000)＝0.8%，阴性预测值＝90 000/(20＋90 000)＝99.98%。

由此可见，由于受患病率的影响，即使试验的特异度很高，当用于患病率很低的人群时，仍

笔记

会出现大量假阳性病人；同样，一种高敏感度试验，当用于患病率很高的人群，仍会出现大量假阴性病人。这就可以解释为什么一项初评诊断价值较高的诊断试验，用于普查时效果并不佳，主要是由于患病率相差过大所致。阳性预测值与敏感度、特异度及患病率之间的关系，可用 Bayes 公式表示：

$$\text{阳性预测值}=\frac{\text{患病率}\times\text{敏感度}}{\text{患病率}\times\text{敏感度}+(1-\text{患病率})(1-\text{特异度})}$$

四、似 然 比

似然比（likelihood ratio, LR）也是可以同时反映敏感度和特异度的复合指标，即有病者得出某一试验结果的概率与无病者得出这一结果可能性的比值。既可计算连续性测量值不同区间的 LR，又可避免简单地将试验结果划分为正常和异常，从而能全面反映诊断试验的诊断价值；同时似然比比敏感度和特异度更稳定，且不受患病率的影响，但不足之处在于该指标是比而不是率，应用时需要在比值与率之间进行换算。

阳性似然比系真阳性率和假阳性率之比，一项诊断试验的阳性似然比为 10，意味着当诊断试验结果为阳性时，判定患病的可能性是不患病可能性的 10 倍。阴性似然比系假阴性率和真阴性率之比，若一项诊断试验的阴性似然比为 0.01，意味着当诊断试验结果为阴性时，判定患病的可能性仅为不患病的 1/100。因此，为肯定诊断之目的，应选择高阳性似然比的诊断试验，为排除某项诊断，则应选择阴性似然比更低的诊断试验。

似然比的应用步骤包括：先通过文献资料估计当前患者的验前概率，最初的验前概率常常是具有某一临床特征（年龄、性别、人群特征）的人群患病率，计算验前比[验前比 = 验前概率 / （1 − 验前概率）]；确定待选择的诊断试验的敏感度、特异度和似然比；再按照诊断试验结果阳性或者阴性估计验后比[验后比 = 验前比 × 似然比]，最终得到验后概率[验后概率 = 验后比 /（1 + 验后比）]。在连续进行多个诊断试验时，前一个试验的验后概率或者验后比就作为后一个试验的验前概率或者验前比。

例如：通过查阅文献获得中国人群乙肝患者 40 岁年龄段肝癌的患病率为 1%。一例 40 岁有慢性乙肝病史的男性患者，前来就诊。如果患者检查发现 AFP 阳性（AFP 诊断肝癌试验的敏感度和特异度分别为 80% 和 90%），阳性结果的似然比为 80%/10% = 8，再按 1% 的验前概率，计算验前比为 1/99，验后比为 8/99，进而该患者的验后概率为 8%，即通过 AFP 检查阳性这一结果，患者肝癌的可能性从 1% 提高到 8%。

若对该患者进一步进行肝脏彩超检查，彩超检查结果也是阳性（发现肝占位并且考虑肝癌），文献报告彩超诊断肝癌的敏感度和特异度都达到 80% 和 80%，阳性结果似然比为 80%/20% = 4。该患者的验前概率 = 8%，验前比 = 8/99，则验后比率 =（8/99）× 4（B 型彩超阳性）= 0.323，验后概率 = 0.323/（1 + 0.323）× 100% = 24.4%，则患者肝癌的可能性从 8% 提高到 24.4%。

若假设不进行彩超检查，而让患者直接接受增强 CT 检查，检查结果也是阳性，即 CT 判断为肝癌，该患者肝癌可能性有多少呢？文献报告增强 CT 诊断肝癌的敏感度为 90%，特异度为 77%，阳性结果似然比为 90%/（1 − 77%）= 3.9，验前概率 = 8%，验前比 = 8/99，验后比率 =（8/99）× 3.9（增强 CT 阳性）= 0.315，那么验后概率 = 0.315/（1 + 0.315）× 100% = 24.0%。

再假如，增强 CT 检查是在彩超之后进行，增强 CT 检查为阳性，该患者患该癌的可能性升高到多少呢？此时的验前比为彩超检查之后的验后比（0.323），而验后比为 0.323 × 3.9 = 1.261，验后概率为 1.261/（1 + 1.261）× 100% = 55.8%，即该患者罹患肝癌的可能性由 24.4% 提高到 55.8%。

因此，在 1% 肝癌患病率基础上，增加 AFP 阳性（阳性似然比 = 8）、彩超阳性（阳性似然比 = 4）、增强 CT 阳性（阳性似然比 = 3.9），患者肝癌患病率提高到 55.8%，如下计算过程：

笔记

验前比 = 1/99，验后比 =(1/99)× 8(AFP 阳性比)× 4(B 型彩超阳性)× 3.9(CT 阳性)= 1.261，则：验后概率 = 1.261/(1 + 1.261)× 100% = 55.8%。

在应用似然比时还要注意以下问题：

(1) 基础患病率不同，诊断试验效率不同。假如该患者不仅为慢性乙型肝炎，影像学及其他检查还发现患者已存在早期肝硬化，其基础肝癌患病率将从 1% 提高到 10%，若系列检查结果显示 AFP 阳性、B 超阳性、增强 CT 阳性，那么该患者肝癌的可能性将达到 93.3%。具体计算过程：验前概率 = 10%，验前比 = 10/90，验后比率 =(10/90)× 8(AFP 阳性比)× 4(B 型彩超阳性)× 3.9(CT 阳性)= 13.9，则验后概率 = 13.9/(1 + 13.9)× 100% = 93.3%。

不同人群的基础患病率不同，运用诊断试验得到的似然比不同，由此估计的验后概率也不同。临床上，确诊疾病需要寻找更多阳性诊断依据，而否定诊断则需要更多阴性结果支持。

(2) 应用似然比应用图直接查找验后概率。如果事先测出一些诊断试验的似然比，依据文献资料获得人群的患病概率(验前概率)，通过计算可推断患者检查后患某病的概率增加或减少了多少，有助于作出正确的诊断。除了运用上述公式进行计算外，还可利用似然比应用图(Fagan's nomogram)(图 9-2)，将直尺的一端放在验前概率对应点，再与该试验似然比所在点对齐，直尺另一端所指就是验后概率，此方法方便易行，临床实用性强。

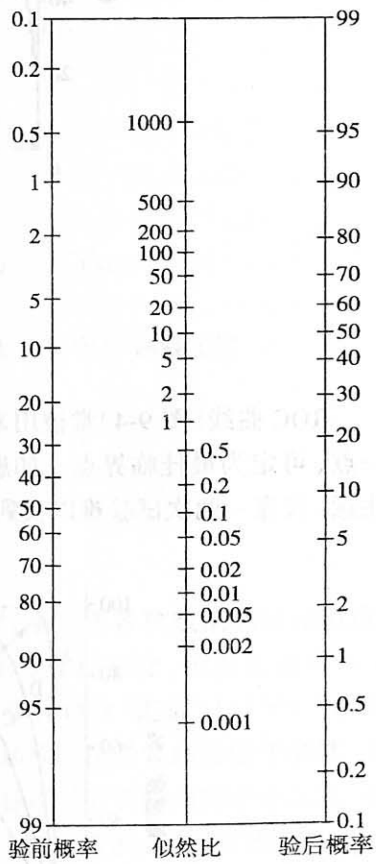


图 9-2　似然比应用图

(3) 当诊断试验结果为二分类变量时，可计算单个阳性或阴性似然比；但当试验结果为连续性变量时，应分别计算不同区间对应的似然比。如试验结果范围是从 0 到 100，则可分别计算诸如 LR(0~10)、LR(11~20)、…、LR(91~100)时的似然比。如：

$$LR(11\sim20)=\frac{\text{病人试验结果在 11\~20 范围内所占概率}}{\text{无病组试验结果在 11\~20 范围内所占概率}}$$

当试验测定结果为连续性变量时，诊断试验特征的最好表达方式是似然比，而不是常用的敏感度和特异度，似然比描述诊断试验的特征更为全面。

第四节　受试者工作特性曲线

前面所列的阳(阴)性似然比、阳(阴)性预测值等指标综合利用了敏感度与特异度的信息，但这些指标都与诊断界点(或阈值)的选取有关。同一诊断试验，不同的诊断界点就对应着不同的敏感度和特异度。为全面评价检测方法的诊断价值，应分别计算不同诊断临界点下的敏感度和特异度。

受试者工作特性曲线(receiver operator characteristic curve，ROC 曲线)，又称受试者工作特征曲线，是以敏感度为纵坐标、以假阳性率(1 - 特异度)为横坐标绘制而成的曲线(图 9-3)。可以进一步计算与比较 ROC 曲线下面积，以综合反映诊断试验的诊断价值。ROC 曲线可以用来：

笔记

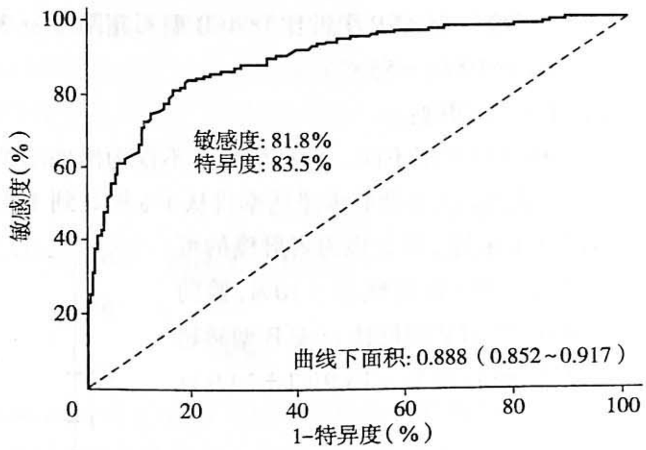


图 9-3 AFP 诊断肝癌 ROC 曲线

一、确定诊断试验的参考值

ROC 曲线（图 9-4）常被用来决定最佳临界点，如患病率接近 50% 左右时，最接近左上角那一点，可定为最佳临界点。如患病率极低或甚高，其最佳临界点可不在最接近左上角那一点。注意：仅靠一两次试验难以找到一个敏感度和特异度俱佳的临界点。

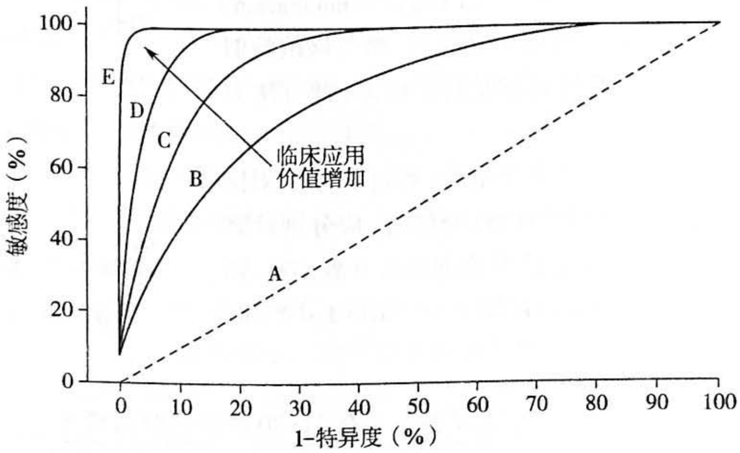


图 9-4 ROC 曲线

二、不同试验诊断价值的比较

ROC 曲线是一种全面、准确评价诊断试验的有效方法，可用来比较两种或多种诊断试验的诊断价值，从而帮助临床医师正确选用诊断试验。图 9-5 中可以看到特异度相同的情况下，血清心肌酶 CK 诊断心肌梗死的敏感度高于心电图 ST 段改变水平的诊断效果。

除了上述目测方法外，还可计算 ROC 曲线下的面积（area under the ROC curve，AUC），来定量比较几种诊断试验的诊断效率，AUC 越大，越接近 1.0，其诊断价值越高，AUC 越接近 0.5，则诊断价值越低。AUC 的计算原理是将曲线下的面积分成多个梯形，计算每个梯形面积，最后将多个梯形面积相加求和，即为 AUC，亦可借助于计算机软件计算。

ROC 曲线具有如下优点：①方法简单、直观，通过目测就可判断和比较诊断价值；②可综合反映敏感度和特异度的相互变化关系；③ ROC 曲线评价与基础患病率无关。但 ROC 曲线同样存在一定的局限性：即 ROC 曲线上所显示的不完全是真正的判断值。

笔记

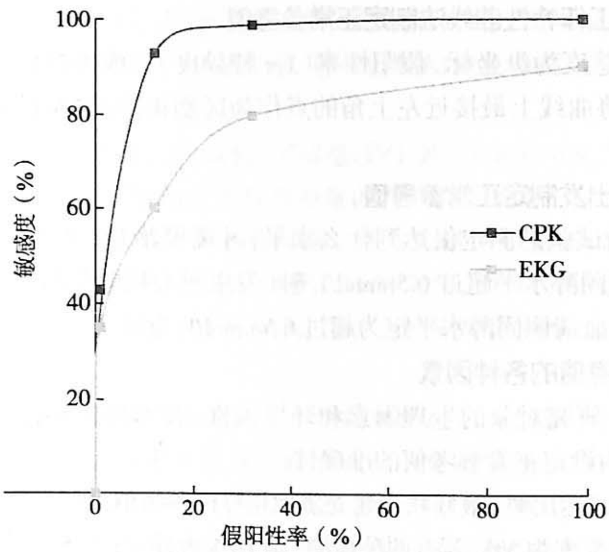


图 9-5　血清心肌酶 CK 水平和 EKG(ST 段改变水平)诊断心梗 ROC 曲线比较

第五节　参考值的确立

一、参考值的概念

医学上对参考值范围的传统概念，指的是正常人解剖、生理、生化等各种数据的波动范围。这些数据不仅因人而异，即使同一个人，还会因机体内外环境的变化而变化，因此需要有一个正常波动范围。传统的认识主要是基于临床实践，着眼于个体，作为划分“正常”与“异常”的界限。随着现代医学的发展，对参考值的概念也有了进一步认识，如预防医学实践着眼于群体，制定不同性别、年龄儿童发育评价标准，制定食品、水、空气的卫生标准及有害物质的允许浓度，作为保护健康的安全界限。这样对参考值的含义就超过了上述狭义的参考值，因而可称为广义的参考值。

二、确定参考值的基本方法

制定参考值，常以“正常人”为对象，这与医学上的健康含义不同，所谓“正常人”不是指机体任何器官、组织的形态及功能都正常的人，而是排除了影响所研究指标的疾病和有关因素后，所确定的同质人群。确定参考值的基本方法如下：

（一）均数加减标准差倍数法

目前在临床上或文献上大多采用“均数 ±2 倍标准差(S)”作为参考值的范围，凡超出此范围的、无论过高或过低均视为异常。采用这种方法确定参考值的前提条件是诊断试验的数据满足正态分布。如有些资料不呈正态分布，可作对数转换变成正态分布后，再计算参考值范围。若转换后仍不满足正态分布，可采用百分位数法。

（二）百分位数法

由于多数诊断试验测定值的频数分布并非正态，因此有人主张用百分位数确定“正常”和“异常”的界限。若诊断指标过高或过低均认为不正常，可采用双侧范围法，确定第 2.5 百分位数($P_{2.5}$)到第 97.5 百分位数($P_{97.5}$)的区间为参考值范围；若数值过小为不正常，则参考值下限定在第 5 百分位数(P_5)；若数值过大为不正常，则参考值上限定在第 95 百分位数(P_{95})。采用百分位数制定参考值范围时，样本量要足够，否则误差比较大，有人建议观察例数至少应在 120 例以上。

笔记

（三）应用受试者工作特性曲线法制定正常参考值

ROC 曲线是用敏感度为纵坐标，假阳性率（1－特异度）为横坐标作图所得的曲线，也可用来确定最佳临界值。将曲线上最接近左上角的点作为区别正常、异常的界限，作为制定正常参考值的依据。

（四）从临床实际出发制定正常参考值

临床上，一个诊断试验的测定值达到什么水平，才需要治疗，常基于人群调查结果进行综合判断。例如，血清胆固醇水平超过 6.5mmol/L 时，发生冠心病的危险性显著升高，而低于此水平危险就不明显，故将血清胆固醇水平定为超过 6.5mmol/L 为异常。

（五）影响正常参考值的各种因素

研究对象的选择、研究对象的生理因素和环境因素、技术操作和仪器设备因素的影响、统计处理问题都可能影响设定正常参考值的准确性。

此外，多项指标同时运用时，最好联合建立多项指标的参考值范围，否则假阳性率会很高。如使用 1 项指标检测的异常率为 5%；若有两项检测，则 10% 为异常；若 5 项检验则 23%[$1-(0.95)^5$]的个体为异常，显然不合理。联合建立多项指标参考值范围的方法可参考医学统计专著。

第六节 多项试验的联合诊断

鉴于敏感度和特异度俱佳的诊断试验不多，所以在不同的临床诊断需求的循证情况下，分别采取多项诊断试验的联合应用，以期提高循证诊断水平。

一、联合诊断方式

联合方式有两种：平行试验（parallel tests）和系列试验（serial tests）。平行试验指的是几个诊断试验同时进行，只要任一试验结果为阳性，联合诊断结果就为阳性。系列试验系按照一定顺序依次进行诊断试验，只有所有试验结果均为阳性者，联合诊断结果方为阳性。因此平行试验可以提高敏感度和阴性预测值，而系列试验相反，提高的是特异度和阳性预测值。

（一）平行试验

平行试验又称并联试验，适用于下列情况：①对于住院或急症病人或外地的门诊病人复诊有困难时，需要迅速作出诊断。②为避免漏诊，但手中只有两项或两项以上不太灵敏的试验时，组合成平行试验特别有用。需要注意的是：平行试验在提高了敏感度和阴性预测值的同时，特异度与阳性预测值会降低，导致假阳性诊断增加。

（二）系列试验

系列试验又称串联试验，适用于下列情况：①为避免误诊情况的发生，可考虑使用系列试验。②当某些试验昂贵且有危险性时，可用较简单安全的试验先做，一旦提示该病可能存在之后，才使用这些试验。如先后使用两种试验诊断疾病，若其余条件均相同，应首先使用特异度高的试验使较少的病人暴露于第二种试验（危险性高），特别是该项试验花费少而又无危险性时，更应先用。③当单项试验的特异度普遍不高时，系列试验也特别有用。

需要注意的是：系列试验可提高特异度和阳性预测值，试验阳性者表明患病更为可信，但同时降低了敏感度和阴性预测值，漏诊的可能性增加。

二、独立性与联合试验的敏感度和特异度

独立性是计算联合试验敏感度和特异度的前提条件：即诊断试验之间是互相独立的，诊断手段和结果相互间没有影响，或者诊断试验的原理截然不同。如诊断肝癌所用的血清学检查 AFP 和影像学检查增强 CT，两者诊断原理不同，就满足独立性。同样，冠心病的诊断，心电图、

笔记

运动试验等反映心肌缺血程度，CTA、冠脉造影反映冠状动脉解剖上的变化，诊断试验间的互相影响也很小；而两种影像学检查如CT和磁共振，都从占位的角度诊断，CT出现占位一般磁共振也可以发现占位，两者结果就存在一定关联。

假设A试验和B试验为两个独立的试验，其敏感度分别为0.8和0.9，特异度依次为0.6和0.9，并假设患病率为20%。联合试验敏感度和特异度的计算方法如下：

（1）系列试验方式：即两者均阳性才算阳性结果。假定先做A试验，A试验阳性者再做B试验。两者均阳性的144例为阳性结果，占总病例数200的72%。即联合敏感度是72%。试验阴性者为480+288=768，占总阴性人数800的96%，即联合特异度为96%。由此可见系列试验提高了特异度。

（2）平行试验方式：即两试验中一个阳性就作为阳性结果。为估算平行试验的敏感度和特异度，先做A试验，A试验阴性者再做B试验。将B试验的真阳性36人加A试验的真阳性160人，共196人，占阳性的200人中的98%，即联合敏感度。这相当于假定A试验和B试验是完全独立的，它们的联合敏感度为0.8+0.9−(0.8×0.9)=0.98。而其阴性432人，占总阴性人数800的54%，即联合特异度。由此可见平行试验增加了联合敏感度。

第七节 诊断试验结果的一致性评价

由于在临床实践或者临床研究中，测量变异（measurement variation）无处不在，它可来自观察者间的变异，观察者的自身变异，测量仪器、试剂的变异及研究对象的生物学变异（个体内及个体间）等，这些变异往往是同时存在、相互叠加的。诊断试验同样如此，那么一项诊断试验结果是否可靠呢？也需要进行评价。诊断试验的可靠性（reliability），又称可重复性（repeatability），是评价诊断试验另一项指标。重复性是指诊断试验在完全相同条件下，进行重复操作获得相同结果的稳定程度。诊断试验结果是否可靠，需要进行一致性评价。

一、Kappa值及其意义

Kappa值是一致性评价的常用指标，用于判断不同观察者间观察结果的一致性。Kappa值的意义和推理举例说明如下：甲、乙两位临床经验相似的医生通过阅读相同的上腹部增强CT片诊断肝癌，共阅读100张CT片，结果表明，两人均诊断肝癌48例，两人均诊断非肝癌38例，观察一致率86%，此临床意见一致率较高，但根据一般常识，任何现象都存在偶然性（机遇）和必然性（非机遇），临床观察也不例外，假如排除了机遇因素后，实际一致率如何呢？实际一致率称Kappa值。Kappa值的计算步骤如下（表9-6）：

表9-6 甲、乙医生阅读同样100张腹部增强CT诊断肝癌的一致率

乙医生	甲医生		合计
	肝癌	非肝癌	
肝癌	48(a)	6(b)	54(r_1)
非肝癌	8(c)	38(d)	46(r_2)
合计	56(c_1)	44(c_2)	100(N)

观察一致率(OA)=$(a+d)/N$=(48+38)/100=86%；机遇一致率(EA)=$(r_1 \times c_1)/N+(r_2 \times c_2)/N$=54×56/100+46×44/100=51%，非机遇一致率=86%−51%=35%，则：Kappa值=实际一致率/非机遇一致率=35%/49%=0.71。

目前对判断Kappa值的意义尚有争议，但多数认为Kappa值在0.4~0.75有中度至高度一致性，>0.75时，一致性好（表9-7）。

笔记

表 9-7 Kappa 值判断标准

Kappa 值	一致性强度	Kappa 值	一致性强度
<0	弱	0.41~0.60	中度
0~0.2	轻	0.61~0.80	高度
0.21~0.40	尚可	0.81~1	最强

二、观察或诊断结果的一致性评价方法

观察结果的一致性评价方法依据数据类型而定，计量资料或连续性变量资料采用组内相关系数(intra-correlation coefficients, ICC)；对于计数资料或分类变量资料，采用观察符合率与Kappa值表示。上述指标大于0.7，可认为一致性较好。

第八节　诊断试验的评价标准

诊断试验研究，在临床应用之前，尤其是用于自己的患者诊断时，需要进行科学性评价。评价的原则包括试验的设计是否合理，测试手段是否可靠，是否避免了偏倚，然后才能对定量指标所示的真实性和可靠性作出精确的判断。

一、是否将诊断试验与标准诊断法(金标准)进行了盲法对比

这是评价诊断试验最核心的一条。作为参照的金标准诊断，定义是否清晰明确，非常关键。在合理选择金标准的同时，待评试验必须同金标准诊断进行独立的盲法比较。即要求试验结果的评价者预先无法获知哪些病例使用金标准判定为"有病"、哪些判定为"无病"，同一病人诊断试验与金标准诊断结果要独立进行评价。

二、研究对象的代表性如何

研究人群应包括两组：一组是用金标准确诊"有病"的病例组，另一组是用金标准证实为"无病"的对照组。病例组应包括各型病例：如典型和不典型，早、中与晚期病例，有无并发症等，以便使诊断试验的结果更具有临床应用的价值。最能体现诊断试验价值的是区分有病变的早期病人和易与该病混淆(症状、体征相同)的其他疾病。因此，诊断试验评价应该纳入那些临床实践中可能遇到的各种病人作为病例组，而对照组应选用金标准证实没有目标疾病的其他病例，特别是与该病易混淆的病例，以明确鉴别诊断的价值，正常人一般不宜纳入对照组，否则会夸大其敏感度和特异度。终末期病人检查时，试验结果常有明显的异常，因此待评诊断试验很容易将晚期病人与正常人区分开来。典型的例子是癌胚抗原(CEA)对结肠癌的诊断价值。最初报道中，36例晚期结肠、直肠癌病人，有35例(97.2%)的CEA升高，对照组为未患结肠癌的其他病人，CEA水平大多较低，据此作者认为CEA是一项筛选结肠癌的有用试验。而之后将诊断试验用于包括早期结肠癌以及有其他胃肠道疾病对象时，发现CEA诊断结肠癌的效率并不高，也不能将早期结肠癌病人与其他胃肠病病人鉴别开来。

三、样本量是否足够

诊断试验中样本量应该足够。过少，样本缺乏代表性；过大又增加工作量和研究费用。

四、参考值选择是否合理

笔记

参考值或截断点(cut-off point)，又称临界值(critical value)，它的选择将直接影响试验的敏

感度和特异度，因此，应交代截断点的选择方法并进一步说明其合理性。

五、是否同时评价了真实性与可靠性

一项好的诊断试验，应该是既真实又可靠的。因此在评价时，不仅要计算反映真实性的指标，同时还要计算评价可靠性的指标，两者缺一不可。

六、是否交代了诊断试验的具体步骤

诊断试验方法（包括所用仪器、试剂、设备、实验条件等）是否具体详尽，操作步骤和注意事项是否明确，以便他人学习或验证，也有助于诊断试验在临床的推广和普及。

七、联合试验的评价

在评价联合诊断试验的诊断价值时，不仅要计算联合试验敏感度、特异度等指标，各个单项诊断试验的评价指标也要一并报告。

八、控 制 偏 倚

评价一项诊断试验时，还要考虑该诊断试验是否排除了各种偏倚（选择偏倚、错误分类偏倚、测量偏倚等）对结果的影响。

九、临床意义及适用性

诊断试验经过效用分析后，还需用可靠的依据说明其临床意义和适用性，包括正确判断的收益和错误判断的可能后果等。

第九节　证据的临床实践及案例

实施诊断试验时，首先要将临床实践中有关诊断试验问题用一个可以回答的方式加以构建，其次证明有关该诊断试验正确性的证据是足够可信的，然后了解证据的重要价值和临床意义。最后考虑如何将这项诊断试验用于当前的病人。具体实践步骤可参照表9-8。

表9-8　有关诊断试验的循证医学实践步骤

1. 根据临床问题（可回答的特殊临床问题）找出最恰当的、高度相关的研究文献
2. 评价诊断试验的科学性
①试验是否与金标准试验进行了独立、“盲法”的比较
②是否每个被测者都用参照试验进行评价
③所研究病人样本是否包括临床中的各种病人
④诊断试验的精确性如何
3. 估计临床应用的重要性
①估计疾病的验前概率
②说明和评估有关试验的敏感度、特异度、预测值和似然比资料
③评估应用该试验似然比估计的期望结果（验后概率）
4. 临床研究结果是否适用于自己的病人并取得预期的结果
①结果是否适用于我自己的病人
②诊断试验结果是否改变了对患病率的估计
③诊断试验结果是否改变了对病人的处理

笔记

一、疾病案例和提出问题

例如：一位35岁男性病人，乙肝病史10年，目前检查HBsAg(+)，HBcAb(+)，HBeAg(+)，ALT正常水平，HBV-DNA检测 5×10^3，血清透明质酸(HA)和Ⅲ型胶原(PⅢP)均不同程度升高，临床怀疑患者肝病程度已进展到早期肝硬化，进一步该如何处理？

首先基于临床案例，提出一个可回答的问题，将问题具体化。如根据现有的临床指南，慢性肝病患者接受抗病毒治疗的标准是不同的，而依据目前患者的资料，是否抗病毒治疗取决于是否进入明显肝纤维化(F3)或者肝硬化(F4)阶段。因此当前需要解决的问题是："患者肝纤维化进展到何种程度？通过何种措施可以帮助我们进一步明确诊断"。

根据一般文献资料或者进展文献，获知肝纤维化的金标准诊断为病理诊断，临床上常用肝穿刺活检病理组织学检查来明确诊断，但为有创性操作，尽管目前通过超声引导穿刺提高了安全性，但很多患者仍难以接受。非创伤性的诊断技术则成为研究和评价热点。血清学检查尽管简便，但临床诊断效率不高。近年来开展了超声弹性成像、磁共振弹性成像在肝纤维化诊断中的应用，那么我们将选择超声弹性成像还是磁共振弹性成像呢？

为此，我们进一步将问题具体化，转换为三个可回答的研究问题是：超声弹性成像诊断肝纤维化有效吗？磁共振弹性成像有效吗？磁共振弹性成像比超声弹性成像更好吗？

二、检索相关文献资料

文献检索通常分三大步：第一，选择医学文献数据库，如最为常用的PubMed(www.pubmed.com)、UpToDate或Ovid等。第二，选择恰当的主题词，如针对本例患者的处理，主题词包括慢性肝病、肝纤维化(fibrosis, chronic liver disease)，超声弹性成像(ultrasound elastography)或者磁共振弹性成像(magnetic resonance elastography)，诊断(diagnosis)。第三，建立检索策略、实施检索。循证医学检索的基本策略是首先选择最新临床实践指南，然后选择指南未纳入的新近meta分析，最后选择未纳入meta分析的原始研究文献。这样可用最少的时间获得最佳证据。

检索式：(hepatic fibrosis)AND(ultrasound elastography OR magnetic resonance elastography [Title/Abstract])。

采用上述检索词，首先检索指南(文献类型：Practice Guideline)，我们在PubMed中发现了2篇实践指南，最新一篇在2011年发表，为西班牙文发表的地方指南，不适合我们现在处理的患者。2010年发表的另一篇意大利地方指南，主要针对非酒精性脂肪肝的诊断与处理，与目前慢性乙肝也不符合。

其次是检索系统综述与meta分析(文献类型：Systematic Reviews)，我们在PubMed上发现16篇文献，2篇与指南重复，1篇关于HCV肝纤维化，1篇关于肝移植术后HCV复发肝纤维化，1篇关于酒精性肝病肝纤维化，1篇为药物性肝损害，1篇非英文文献，1篇与主题无关；6篇为2007—2010年期间发表的meta分析，其中1篇2011年发表在*Journal of Hepatology*上的超声弹性成像meta分析和1篇2012年最新发表在*Journal of Hepatology*上的磁共振弹性成像符合我们的检索要求。但未发现比较超声弹性成像和磁共振弹性成像的诊断研究，需要在原始研究文献中继续检索。

最后，检索原始研究文献。首先，如果我们采用上述检索式检索全部杂志，有504篇之多，若将检索式中超声弹性成像或者磁共振弹性成像中的OR修改为AND，目标将缩小到19篇。从中我们很快找到2008年发表在*Gastroenterology*中的1篇原始研究文献，但已被两篇meta分析纳入，meta分析发表之后未发现更新的原始研究文献。

笔记

三、评价检索资料的科学性

当收集到了相关的文献资料后必须按照表 9-8 中第 2 点提出的标准对文献进行严格评价。通过者才可认为：设计合理、科学，其结果真实可信。以上述原始研究为例进行阐述。

（一）试验是否与金标准试验进行了独立、“盲法”的比较

判断诊断试验真实性的最好方法，是将所考核的诊断试验结果与“真实”情况进行比较。真实情况是由金标准诊断来确定的。金标准的选择应结合临床具体情况而定，因此，当找到相关文章后，首先要检查每一位研究对象在进行诊断试验时，是否都采用了合适的金标准。前文中所举的例子，其金标准是病理组织学检查或者通过肝穿刺活检，若腹水或者凝血功能等影响肝穿刺，改用经颈静脉肝静脉活检获得组织学诊断。待评诊断试验为超声弹性成像和磁共振弹性成像，在方法学中都进行了描述，每种检查都是独立进行的，结果评价者不知道最终组织学诊断结果，也不知道另一项检查结果。

（二）是否每个被测者都用参照试验进行了评价

有些情况下，如金标准昂贵或具侵入性，可能不是所有的患者都进行了检测。研究者常常将试验结果阳性者，送去做金标准检测，而阴性者只抽一部分人去做，这样必然夸大了诊断试验的敏感度，造成偏倚。在上述例子中，对没有完成金标准诊断的 13 例患者进行了剔除，这样的报告更客观和真实。

（三）所研究病人样本是否包括临床中的各种病人

上述例子中包括了临床上从无肝纤维化到明显肝硬化的各种病人，即便存在腹水也通过特殊检查获得组织学诊断纳入。

同时也应看到，文中部分患者由于种种原因，造成超声弹性成像（腹水 13 例，肥胖 10 例）或者磁共振弹性成像（幽闭恐惧症 3 例，血色病 3 例，肥胖 2 例）检查结果无法判定或者评价失败。这些患者实际就是将来诊断试验应用时，应限制的人群，在个体化选择诊断试验时将起决定作用。

（四）诊断试验的可靠性

上述例子中，作者虽没有描述变异系数和 Kappa 值，但全部测量结果都经过了前后两次评价，并且至少有两名高年资或者操作有丰富经验的专业人员评价，每种方法的评价都给出了具体的标准。在其他研究报道中两种诊断试验结果测量，也是可重复的。

如果表 9-8 中第 2 点标准中有至少一项不达标，就说明诊断试验可能存在严重缺陷，使用时应慎重。但要找到完全符合上述标准的诊断试验文献有一定的难度，为此，一些国际组织或学术机构开始出台一些临床研究报告规范，以促进研究质量和研究水平的提高。其中，对于诊断试验研究，2003 年初在 *Clinical Chemistry*, *Annals of Internal Medicine* 和 *BMJ* 等主流杂志上发布了诊断试验准确研究报告标准（Standards for Reporting Diagnostic Accuracy, STARD），为评定诊断试验提供了详细的可遵照标准。STARD 包括 25 个条目，并采用了图标设计以提高评价诊断研究报告准确性的方法学和信息质量。

四、实践步骤

（一）估计诊断试验临床应用的指标

1. 估计疾病的验前概率　上述案例中提示早期肝硬化的特征包括：35 岁，10 年慢性乙肝病程、血清学 HBsAg(+), HBcAb(+), HBeAg(+), 血清肝纤维化指标阳性。临床判断早期肝硬化的概率在 50% 以上。相反，如果 25 岁患者，仅仅体检发现 HBsAg(+), 其他病毒标志物阴性，肝纤维化指标均阴性，我们估计早期肝硬化的概率不足 1%。即使年龄 35 岁患者，母亲有乙肝携带状态，并且患者 HBsAg(+)来源于其母亲，早期肝硬化的概率也不超过 10%。

笔记

由于患病率与诊断试验的阳性预测值(试验结果阳性时,患者患病的概率,即验后概率)成正比,与诊断试验的阴性预测值(试验结果阴性时,患者不患该病的概率)成反比。正确地估计验前概率就显得十分重要。如果验前概率的估计来自他人报告的结果,应考虑自己病人的情况是否与其报告的一致,若不同,则需查阅更多文献以了解不同情况下的验前概率。

2. 说明和应用关于诊断试验结果的资料 诊断试验评价指标包括敏感度、特异度、预测值、ROC 曲线下面积、阳性和阴性结果似然比等指标,反映试验的诊断效率。通常特异度高的试验用于肯定疾病诊断。敏感度高的试验用于疾病筛查。相对于单个试验,meta 分析资料更有利于临床实践评估(表 9-9)。

表 9-9 磁共振弹性成像和超声弹性成像诊断明显肝纤维化和早期肝硬化的价值

	磁共振弹性成像	超声弹性成像	*P* 值
明显肝纤维化(F0~F1)vs.(F2~F4)			
敏感度	0.94(0.81~0.98)	0.79(0.74~0.82)	
特异度	0.95(0.87~0.98)	0.78(0.72~0.83)	
SROC	0.994(0.985~1.0)	0.837(0.756~0.918)	<0.05
早期肝硬化(F0~F2)vs.(F3~F4)			
敏感度	0.92(0.85~0.96)	0.83(0.79~0.86)	
特异度	0.96(0.91~0.98)	0.89(0.87~0.91)	
SROC	0.985(0.968~1.0)	0.906(0.838~0.975)	<0.05

3. 应用似然比 似然比的含义是试验的结果使验前概率提高或降低了多少。实际为验前比提高或者降低的倍数。表 9-10 为基于验前概率计算的验后概率。

表 9-10 磁共振弹性成像检查判断明显肝纤维化存在与否计算的验后概率

验前概率	验前比	试验结果	似然比	验后比	验后概率
50%	1	+	18.8	18.8	94.9%
50%	1	−	0.063	0.063	5.9%
10%	1/9	+	18.8	2.09	67.6%
10%	1/9	−	0.063	0.007	0.7%
1%	1/99	+	18.8	19.0	16.0%
1%	1/99	−	0.063	0.0006	0.06%

(二)将临床研究结果用于自己的病人

如何将经过严格评价后的诊断试验结果,应用在我们遇到的具体病人身上,应从以下三个方面去考虑:

1. 试验结果是否适用于我的病人 首先要确定在本单位或者推荐的检查单位是否已开展或能开展该项检查,包括仪器、设备、试剂、人员的配备。还要考虑将该试验搬过来用于自己病人的测定时,一致性和重复性如何。其次,对比系统综述或文献报道的研究结果所报道的研究场所(如美国还是中国)、其文化背景、人口特征等是否与我所在单位病人的情况相同。两者的情况越接近,试验结果的重复性就越好。在本例中,肥胖患者的诊断无论超声弹性成像或者磁共振弹性成像都存在检查结果无法判定的问题,肥胖对超声弹性成像的影响更大。超声弹性成像同时受腹水影响,但对磁共振没有影响。

2. 如何合理估计自己病人的验前概率 可以根据个人临床经验、地区性和全国性资料或文献所报告的人群调查结果进行估计。对于一篇报道某一疾病患病率的文献可用如下标准进

笔记

行评估：

（1）有关验前概率的证据是否科学？具体包括：①被研究病人是否代表了有这个临床问题的完整的疾病谱？②对于确诊的每一条标准是否明确和可靠？③诊断的程序是否全面并能持续应用？④对于最初没有诊断出来的病人是否进行了足够长时间和完整的随访？

（2）有关验前概率的证据是否重要？具体包括：①诊断的是什么病？其概率是多少？②对于疾病概率估计的精确程度。

符合上述标准文献数据可以用于自己病人患病概率的估计。分两种情况：若诊治条件和病人特征与文献类似，可直接采用文献的验前概率；若诊断条件、病人特征与文献报道不同，则可以文献报告的验前概率作为基点，根据病人实际情况进行调整，观察验后概率的变化，最后确定验前概率。

3. 验后概率的结果是否改变了对病人的处理　在本例中，35 岁，10 年慢性乙肝病程、血清学 HBsAg(+)、HBcAb(+)，HBeAg(+)，血清肝纤维化指标阳性。临床判断早期肝硬化的概率 50%，即验前概率为 50%，诊断试验阳性，肝纤维化或者肝硬化验后概率达到 94.9%，抗病毒治疗就非常必要，反之，若概率低于 5.9%，可暂时不选择抗病毒治疗。

若一名 25 岁就诊者，仅仅体检发现 HBsAg(+)，其他病毒标志物阴性，肝纤维化指标均阴性，我们估计早期肝硬化的概率不足 1%。即使磁共振弹性成像诊断试验结果为阳性，验后概率仅 16%，抗病毒治疗的证据并不充分。

又假设一名年龄 35 岁患者前来就诊，母亲有乙肝携带状态，并且患者 HBsAg(+)来源于其母亲，早期肝硬化的概率估计 10%，磁共振弹性成像结果阳性，验后概率 67.6%，如何选择处理，是抗病毒治疗或者暂时不抗病毒治疗？

这里引出了“诊断阈值”和“治疗阈值”两个概念。决定一项诊断试验是否重要时，我们关注的是该试验在鉴别有病和无病时的正确性以及其改变我们在做试验前估计到病人患病可能性与试验后估计的患病可能性的能力，能较大地改变验前和验后概率的试验，对我们临床实践是重要和有用的。也就是说只有当某一项诊断试验能使验前和验后概率发生较大变化，而且这种变化会对是否要继续进行另一项检查或对治疗计划的改变有影响时，才能认为这项诊断试验的确对病人是有帮助的。

如果诊断试验的目的是为了决定是否对病人进行治疗，此时不仅要了解该诊断试验的有效性，更重要的是评估通过该项试验的阳性或者阴性结果带来的对病人治疗方案决策的结果是否发生变化。新试验是否增加了信息，如果做了该试验后，能增加已有的信息导致治疗措施的改变，从而最后给病人带来益处，这项试验就是有价值的。因此，试验原则是使用有可能改变处理方案的试验，而治疗原则是选用利大于弊的治疗方案。我们用早期肝癌诊断与处理为例说明行动点、诊断阈值和治疗阈值。

让我们先来看看治疗的情况。早期肝癌的治疗，手术切除作为首选，但即使这样，手术仍存在一定风险。我们假设早期肝癌手术后 3 年存活率为 80%，疾病本身、肝功能状态以及手术创伤的死亡率达到 20%，而如果采取其他措施包括积极观察病情变化而不立即采用手术治疗，可能的结果是：存活率下降到 50%，死亡率上升到 50%。与暂时观察相比，及时手术显然利大于弊：80% − 50% = 30%，即存活率提高 30%，死亡率下降 30%。如果没有别的选择，合理的决策一定是选择及时手术治疗（表 9-11）。

表 9-11　对于已确诊的疾病

是否有病	是否治疗	存活率	死亡率
有	是	80%	20%
有	否	50%	50%

笔记

上面的处理决策是在疾病诊断明确的情况下，手术方案利大于弊，有高于30%的存活率，但假如疾病诊断的概率不是100%呢？出现诊断错误时，如果选择暂时观察，不会出现手术带来的风险，即使需要择期手术，也几乎可完全治愈；相反，如果选择及时手术，在病人一般状况没有改善的情况下，手术本身的风险及死亡率可能达到10%。因此，与暂时观察相比，及时手术的结果是增加了10%的死亡风险。对不能确诊的疾病，临床决策还依赖于对疾病诊断的可能性大小，即诊断概率（表9-12）。

表9-12　对于尚未确诊的疾病

是否有病	是否治疗	存活率	死亡率
有	是	80%	20%
有	否	50%	50%
无	否	100%	0%
无	是	90%	10%

如果诊断该疾病的可能性，即概率为90%，选择及时手术治疗：患者得到的好处为27%，即90%（患病可能性）×（80%－50%）（生存率增加），而可能出现的害处是1%，即10%（诊断错误可能性）×（100%－90%）（生存率减少），此时的合理决策仍是选择及时手术。

如果诊断疾病的可能性，即概率为10%，选择及时手术治疗：患者得到的好处是3%，即10%×（80%－50%），而可能出现的害处是9%，即90%×（100%－90%），合理决策是暂时不考虑手术。

在10%~90%之间，可以找到一个概率点，在这点上，治疗的利弊相当。在这点之上，治疗利大于弊，选择该治疗；在这点之下，治疗弊大于利，暂时不选择该治疗。在本例中，患病可能性×（80%－50%）=（1－患病可能性）×（100%－90%），患病的可能性，即行动点为25%。

在上述的案例分析中，前提条件是没有可以帮助进一步确定或者排除该疾病的诊断手段。因此，如果患者不是来源于肝硬化人群，也没有B型彩超结果阳性，依据患者有慢性乙型肝炎基础（基础概率1%），AFP阳性（阳性似然比＝8），增强CT检查阳性（阳性似然比＝3.9），该患者肝癌可能性达到24.0%，如果没有其他试验可以进一步选择，可以直接决策，即选择手术切除治疗。

而实际上，肝癌的诊断还有很多其他方法。一方面，通过检查明确患者是否有肝硬化，增加肝癌患病可能性（即基础患病率），另一方面，接受如磁共振检查、血管造影检查、超声引导下穿刺活检病理组织学检查等帮助我们进一步确定或者排除肝癌。治疗决策又会是怎样？是否需要在术前选择其中一种或者多种检查？

处理（手术）与不处理（暂时不手术）的利弊已经明确，行动点在25%的概率水平。目前患者肝癌的可能性估计为24.0%。如有更多诊断试验可以选择，是否选择取决于诊断试验的优劣。

首先考虑磁共振，除了增加费用之外，没有过敏等副作用，安全性很好，医院可以提供（可获得性），患者接受程度也很高（可接受程度），是否选择、取决于其诊断效率，即敏感度和特异度。文献报告磁共振诊断肝癌的敏感度为86%，特异度为81%。目前患肝癌诊断可能性仍为24.0%，试验结果阴性，疾病的可能性，即概率会减少，没有改变随访选择。但若试验结果阳性，疾病的可能性，即概率增加，从24.0%提高到25%之上，则可能支持及时手术治疗。因此我们选择磁共振检查。

在上述案例中，如果患者B超和CT都发现占位考虑肝癌，患病可能性为55.8%，我们是否还会选择进行MRI检查呢？试验结果阳性，直接手术，没有改变我们的决定，无须检查；试验结果阴性，根据试验的敏感度和特异度，推算：

诊断试验的阴性似然比为：（1－敏感度）/特异度＝（1－86%）/81%＝14/81

笔记

验前概率为 55.8%，验前比为：55.8%/44.2%＝1.26；验后比为：1.26×（14/81）＝0.218，则：验后概率＝0.218/（1＋0.218）×100%＝17.5%。

这一结果已低于行动点（概率为 25%），所以应该为患者选择磁共振检查，因为检查结果阳性更支持肝癌诊断，支持手术治疗，而阴性结果则可能改变我们手术治疗的选择。

另一种情况，假如我们知道，该患者基础患病的可能性为 1%，是否直接接受磁共振检查（如作为一种高危人群普查项目）而不先进行 AFP 检查、B 型超声检查、CT 检查呢？让我们先分析一下直接检查磁共振的结果。如果磁共振试验阴性，患病可能性即概率减少，不支持及时手术治疗，没有改变选择；试验结果磁共振阳性，疾病的可能性即概率会增加。根据试验的敏感度和特异度，阳性似然比为：敏感度 /（1－特异度）=（86%）/（1－81%）＝4.5；验前概率为 1%，验前比为：1%/99%＝1/99；验后比为：（1/99）×4.5＝4.5/99＝0.046。

则：验后概率 =（0.046）/（1＋0.046）×100%＝4.4%。

这一结果没有高于行动点（概率为 25%），无论该试验结果如何，仍然选择放弃手术治疗，诊断试验对治疗决策没有影响，因此直接选择该诊断试验是不合理的。

如果从 1% 基础水平逐渐增加患病的可能性，可以找到一点，在这点之下，无须进行该诊断试验，因为无论该试验结果如何，我们都选择观察，在这点之上，进行诊断试验，如果试验结果阳性，患病的概率可能上升到 25% 以上，将改变处理患者的决策，这一点称为诊断阈值（test thresholds）。

诊断阈值比 = 行动点比 / 阳性似然比 =（1/3）/［（86%）/（1－81%）］= 1/45＝0.074，则：诊断阈值概率＝0.074/（1＋0.074）×100%＝6.9%。

从 90% 逐渐往下降低患病概率，可以找到一点，在这点的患病概率上，无须进行该诊断试验（磁共振检查），因为无论试验结果如何，我们都选择手术治疗；在这点之下，进行诊断试验，如果试验结果阴性，患病的概率下降到 25% 以下，将改变我们处理患者的决策，这一点称为治疗阈值（treatment thresholds）。

治疗阈值比 = 行动点比 / 阴性似然比 =（1/3）/［（1－86%）/81%］= 1.69，则：治疗阈值概率 = 1.69/（1＋1.69）×100%＝62.7%。

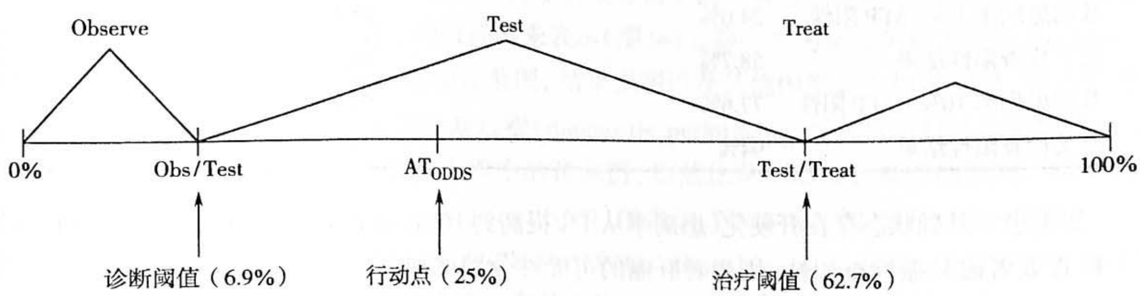


图 9-6　行动点、诊断阈值、治疗阈值示意图

如果诊断概率出现下列情况，以下为合理处理：①诊断阈值之下，无须进一步诊断试验，直接决定暂时观察；②介于诊断阈值与行动点之间，需要进一步诊断；若没有其他诊断试验，这时采取的治疗害处大于好处，采取治疗要慎重；③介于行动点与治疗阈值之间，需要进一步诊断；若没有其他诊断试验，这时采取的治疗好处大于害处，可以采取治疗；④大于治疗阈值，无须进一步诊断试验，直接决定治疗；⑤介于诊断阈值与治疗阈值之间，如有其他诊断试验，需要进一步诊断，直到达到治疗阈值才开始治疗（理论上）。

可以看到，患者在基础状态下，肝癌患病率仅 1% 时，我们首先选择了 AFP 检查与 B 型超声检查，因为在肝癌患病率低于 6.9% 的人群中直接进行磁共振检查即使阳性也不适合直接选择手术处理。而直接选择磁共振检查也违背了诊疗流程遵循的其他原则包括检查费用低、安全性

笔记

好、可接受程度高、可获得性高(在一般医院都可以获得)等。

回到我们的临床案例中,患者AFP阳性结果,肝癌患病率从1%提高到8%,我们在前面的案例中分析了增强CT检查,并进行了选择获得了阳性结果。我们是否可以不选择增强CT检查而直接选择磁共振检查呢?

如果直接选择磁共振检查,阳性结果似然比为4.5,患者阳性结果,肝癌患病概率从8%提高到28.1%,超过决策处理的25%,因此同样可以选择。让我们比较以下两种试验的差异(表9-13),由于磁共振的特异度更高,而且无造影剂过敏或者造影剂可能带来不良影响的疾病如肾功能不全,可以优先选择磁共振。阳性结果提供了更高的患者肝癌诊断可能性(24.0% vs. 28.1%)。但无论首先选择何种试验,阴性结果可以就此观察,而阳性结果都需要进一步选择另一个试验加以验证。

表9-13 增强CT与磁共振诊断肝癌检查比较

比较内容	增强CT检查	磁共振检查
诊断基础	影像学特征,占位	影像学特征,占位
诊断效率		
敏感度	90%	86%
特异度	77%	81%
阳性结果似然比	3.9	4.5
阴性结果似然比	10/77	14/81
诊断试验特征		
可获得性	三级医院	三级医院
易接受程度	中等(相比抽血/造影或活检)	中等(相比抽血/造影或活检)
费用(元)	600~1000	600~1000
安全性	有造影剂过敏可能,可能带来肾脏影响	过敏可能小,体内存在金属物不适用
在该患者中的应用		
基础患病率1%+AFP阳性	24.0%	28.1%
交叉检查阳性结果	58.7%	58.7%
基础患病率10%+AFP阳性	77.6%	80.0%
交叉检查阳性结果	94%	94%

如果患者基础状态存在肝硬化(患病率从1%提高到10%,AFP阳性,患病率提高到30.7%),CT检查或者磁共振检查阳性,则患者肝癌的可能性分别增加到93%和87%,都不需要进一步选择另一项检查印证,而可以直接处理。可见基础患病率即来源人群不同明显影响诊断试验对结果的判定作用。

是否CT阳性肝癌诊断达到77.6%的可能性或者磁共振阳性肝癌诊断达到80%的可能性就一定能选择手术治疗呢?或者CT与磁共振检查其中一项为阳性结果,另一项为阴性结果,我们应该考虑进一步检查如血管造影检查(表9-14)。其敏感度和特异度均达到98%和98%,阳性结果似然比为49,阴性结果似然比为1/49。在很多情况下血管造影或者穿刺组织学检查作为评价CT或者磁共振等检查的金标准。但因其具有一定创伤性、部分患者难于接受、存在一定风险并且费用较高等,只有在存在争议、各种检查不一致的情况下临床才选择。

敏感度=98%,特异度=98%,阳性预测值=98/100×100%=98%,阴性预测值=98/100×100%=98%,阳性似然比=98%/2%=49,阴性似然比=2%/98%=1/49。

笔记

假设CT检查结果阳性,患者来源于肝硬化人群,肝癌可能性77.6%。如果血管造影结果阳

表 9-14 血管造影诊断肝癌临床评价

	金标准评价最终结果	
	肝癌	非肝癌
血管造影判断(+)	98(98%)	2(2%)
血管造影判断(-)	2(2%)	98(98%)

性，疾病的可能性即概率增加，支持及时手术治疗，没有改变选择；如果试验结果阴性，疾病的可能性即概率会减少。根据试验的敏感度和特异度，推算验后概率为 6.6%，低于 25%，仍应该选择进行这项检查。

阴性似然比为：(1－敏感度)/特异度＝(1－98%)/98%＝1/49；验前概率为 77.6%，验前比为：77.6%/22.4%＝3.46；验后比为：3.46×(1/49)＝0.071；则：验后概率＝0.071/(1＋0.071)×100%＝6.6%。

特别需要强调的是，权衡利弊还需要考虑决策时间，利弊可以用临床疗效和不良反应评估，也可以包括生存质量、临床成本。如果我们知道抗病毒治疗与暂时不抗病毒治疗的利弊，我们就可以回答前面案例中提及的患者，通过磁共振弹性成像诊断早期肝硬化验后概率上升到 67.6%，是否需要选择抗病毒处理，或者是否有必要进一步采用创伤性检查即肝穿刺活检评估诊断。临床上疾病常常处于动态变化过程中，临床疾病诊断的概率也在不断变化，对于不能等待的疾病或者状态，需要及时作出决策；对于可以等待的疾病或者状态，可以选择暂时观察，但需要评估等待的风险。

五、小 结

总之，我们在开出医嘱做某项诊断试验前应考虑：①验前概率是多少？对病人诊断还有多大疑问？是否需要做这项检查？②该项检查如果漏诊或误诊会对病人带来多大危害？③这项检查的似然比能否改变进一步临床决策。当然还应考虑到做该项检查的危险性、费用以及做该项检查的迫切性。通过实施上述步骤使医生对病人的处理更具科学性。

Price 曾提出诊断试验的结果常常被作为是决策分析中的一部分，从实施诊断试验开始到最终作决策之间的各阶段以自下到上的阶梯式来表示(图 9-7)，第一步为试验的技术指标(technical performance)，包括试验的精确性、准确性范围，结果说明以及其他因素如生物学变异、标本的稳定性对试验结果的影响。第二步为诊断过程(diagnostic performance)，包括对目标人群的测定，试验的敏感度和特异度、不同患病率人群中的预测值，似然比及其对验后概率的影响。也可用需要诊断的人数(number needed to diagnose，NND)来表示，NND＝1/[敏感度－(1－特异度)]。NND 有助于试验之间的比较。第三步为对临床的影响(clinical impact)，包括：①对诊断步骤的影响，例如与其他试验比较，对提高诊断效率的作用；②对治疗策略的影响，例如对选择最优治疗手段、减少和避免治疗的副作用方面有何影响；③对健康结局的影响，例如微球蛋白测定对早期发现糖尿病肾病、较好地处理糖尿病的伴发病高血压以及降低肾功能衰竭方面的作用。第四步为对组织管理的影响(organization impact)，包括应用某诊断试验后是否可减少住院天数、减少医生所花费的时间和设备的消耗、是否降低了医疗资源的消耗。第五步是经济评价(cost-effectiveness)，通过成本 - 效果分析，了解诊断试验的经济效果。特别当一项新试验比老的试验价格贵，但诊断效果高，此时不能仅将注意力集中在试验本身，而要测定病人的结局以及通过经济评价得出结论。总之，以上每一步对决策分析均十分重要，根据上述各步提供的证据，再结合自己病人情况才能作出是否要采用该项试验的决定。

决策的最终结果是为了改善病人的健康结局。如果一项试验可以改变诊断步骤或(和)治疗策略，其结果在总体上可改善病人健康结局的，就是对临床有益的试验。

笔记



图 9-7 设计利于决策分析的循证实践证据阶梯

目前循证医学在诊断试验范围内的应用还不够广泛，突出表现在不同医院对同一种疾病采用的诊断步骤和检查项目各不相同，差异很大，其中有许多属于诊断试验应用不当或选用不合理。因此，在具体病例的实践中，要学会选择恰当的诊断试验，并对患者做出循证决策。

（陈世耀　张宁萍）

第十章　疾病治疗的研究、评价与循证实践

临床治疗研究与评价是临床流行病学与循证医学的重要内容之一。目前，许多疾病的治疗药物、方法和手段日趋多样化，如何从中选择安全、有效的干预措施，指导临床实践决策，已成为临床工作的一项重要任务。努力学习和掌握必要的治疗性研究与评价方法对医务工作者具有重要的意义。

治疗性研究是指在临床实践中以人为研究对象，应用医学科研的理论和方法，通过严谨的设计（design）、精确的测量（measurement）对欲选择的治疗方法或方案进行的科学研究与客观评价（evaluation）。高质量的研究结果可以作为证据，再进一步结合患者意愿和现有医疗条件制定出最佳治疗决策，以达到提高治愈率，降低病残及病死率，提高生存质量以及改善人体健康的目的。

治疗性研究与评价的内容既可以是药物、术式等治疗手段，也可以是各种预防性干预措施。既可以进行原始研究与评价，也可以进行二次研究与评价。

第一节　治疗性研究与评价概述

一、治疗性研究与评价的重要性

在临床实践中，常常需要观察和评价某些措施疗效的大小。但由于存在可能影响疗效评价的其他因素，例如样本量大小、样本代表性及组间可比性等，因此研究结果不一定是真实可靠的。所以对那些自称具有"显著"临床疗效的治疗措施，只有经过科学研究和严格评价才能确认。

首先，治疗措施在体外实验中观察到的效果须经严格的临床验证才能证明其是否有效。以动物为研究对象或体外实验完成的结论不能直接用于人体。如体外实验证实阿糖胞苷（Ara-c）有干扰嘧啶合成的作用，可抑制多种 DNA 病毒，能抑制播散性带状疱疹病毒，但经临床试验发现，Ara-c 不仅对播散性带状疱疹的临床治疗无效，反而有害。

其次，临床试验过程中，特别是在设计、资料收集、分析评价等方面均可能受偏倚的影响，得出错误的结论。因此，只有经过科学验证和严格评价，才能确定新药物、新疗法的真实效果。

最后，临床经验本身不是科学证据，也只有经过科学的效果评价，才能确认。如：曾风行一时的胃冰冻疗法治疗消化性溃疡、乳房内动脉结扎治疗顽固性心绞痛以及国内曾盛行的鸡血疗法、甩手疗法、针灸哑门穴治疗聋哑等，后经科学证实均无效。

近些年来，随着临床流行病与循证医学的发展，许多临床医生已经逐渐地认识到应用科学的研究设计，如多中心、前瞻性、随机双盲、平行对照临床试验对临床疗效评价的重要性。开展高质量的临床试验，既能保证获得真实可靠的结果，又会促进临床医学研究和临床工作的不断发展。

二、治疗性研究与评价的基本条件

一项治疗性研究能否得到实施，一般要考虑下列条件：

（一）具有充分的科学依据

每项临床疗效评价研究都要有清楚的研究背景和充分的立题依据，以及明确的试验目的，而不能仅凭经验来做假设或推断。早在 1964 年第 18 届世界医学大会通过并经数次修改的《赫尔辛基宣言》（*The Declaration of Helsinki*），是人体试验应遵循的基本准则，规定："凡涉及人体

笔记

的生物医学试验，必须遵循科学的原则，应建立在足够的实验室和动物实验及科学文献认识的基础之上”。意味着用于人体治疗性试验的任何药物或措施，在针对疾病发病的机制或对某种环节的干预应有理论依据，同时还应该具有药物化学、药理学、毒理学、药动学以及药效学等研究基础。在初步证明具有安全性和有效性之后，方可投入临床治疗性试验。如单硝酸 -s- 异山梨酯醇（isosorbide-s-mononitrate，IS-S-MN）对预防性治疗食管静脉再出血的研究，研究者在研究前首先详细地了解了食管静脉曲张出血及再出血的发生率和病死率，同时了解了国内引起食管静脉曲张出血的主要原因，进而分析了食管静脉曲张出血的各种影响因素，比较了内镜和普萘洛尔等治疗方法的优缺点等。通过反复论证，研究者提出 IS-S-MN 可能比普萘洛尔的副作用小但疗效佳的假设，尽管研究者发现国外已有此方面的文献报道，但样本量太小，尚不能作出肯定结论，故认为有必要进行此项临床治疗试验。

（二）明确研究设计拟达到的最佳目标

不同疾病均具有各自特点和不同的结局表现，同时药物或治疗措施也都有着各自不同的治疗效能。利用某种药物或某种治疗措施达到疾病根治的目的往往不现实，因此，治疗的预期目标，除了要考虑治愈和根治外，还要考虑症状缓解，功能维持，预防复发或并发症等，应基于疾病的具体特点和试验药物本身的效能进行审慎的选择。如果所观察的疾病是可被治愈或根治的，那么治愈或根治就为研究拟达到的最佳目标。如外科手术可达到根治早期肿瘤的目的等；而有些疾病在某种情况下有可能复发或发生某种并发症而引起更为严重的后果，那么有效地预防复发或减少某些并发症的发生，就作为拟达到的目的。如对脑梗死恢复期病人，临床治疗的目标应是预防其再次复发；而某些不能彻底治愈的慢性病患者，临床治疗的目标主要在于缓解症状，最大限度地改善其功能状态及生存质量。如类风湿疾病治疗等，其目的就是控制疼痛、防止畸形、改善生存质量等。

（三）选择最有效的试验药物或措施

临床治疗性药物或措施的选择，除了要有科学依据及安全有效性证据外，还应考虑从同类药物或措施中选择最为有效者进行试验，避免低效或无价值的重复试验。此外，尽管有些药物或措施可能不会显著提高疗效，但其副作用发生率或者成本下降，也应优先考虑进行试验。

（四）选择最佳治疗水平的终点指标

评价治疗效果要选择合适指标，这些指标包括终点指标和中间指标。终点指标代表疾病的最终结局，如某些肿瘤外科手术或化疗后不同时期的存活率或死亡率；中间指标是指疾病发展变化过程中的一些结果，如药物治疗乙型肝炎期间转氨酶的变化水平及其他实验室指标水平等。治疗终点指标的选择应基于疾病的性质、病损程度、治疗后机体的病理损害和生理功能状况的可复性而定。例如：治疗高血压病的目标是使血压下降到最佳治疗水平、防止心脑肾等靶器官损害事件发生。

（五）明确研究中伦理问题

在临床研究中，所有以人为对象的研究必须符合世界医学会《赫尔辛基宣言》、国际医学科学组织理事会《涉及人的生物医学研究国际伦理准则》的相应规定，即公正、尊重人格、力求使受试者最大程度受益和尽可能避免伤害发生。

从药物临床研究设计阶段就要重点考虑伦理问题。考虑以下 8 方面内容：即方案设计与实施、试验的风险与受益、受试者的招募、知情同意书告知的信息、知情同意的过程、受试者的医疗和保护、隐私和保密、涉及弱势群体的研究等。

例如在开始一种新药的 RCT 之前，应该明确 RCT 的设计和执行方案是否科学？是否考虑受试对象的收益和风险比？是否确保受试者已经知情同意而且是自愿参与？是否可以随时、随地地自由退出试验？是否能保护受试者的秘密？是否能迅速发现和处理不良事件并给予相应的补偿或者赔偿？一旦有更好的治疗方案，是否能停止 RCT？

笔记

三、治疗性研究与新药临床试验分期

治疗性研究中，最常见的研究方案是临床试验，其中又以新药临床试验居多。按照新药的不同研发阶段，药物临床试验进一步可分为：

（一）Ⅰ期临床试验

Ⅰ期临床试验为初步的临床药理学及人体安全性评价，是在大量实验室研究、试管实验与动物实验基础上，将新疗法开始用于人类的试验，是在人体进行新药试验的起始期。目的在于了解剂量反应与毒性，进行初步的安全性评价，研究人体对新药的耐受性及药代动力学，了解药物在人体内的吸收、分布、消除的规律，以提供初步的给药方案。受试对象一般为健康志愿者，在特殊情况下也选择病人作为受试对象。方法为开放、基线对照、随机和盲法。Ⅰ期临床试验的样本量为20~30例。

（二）Ⅱ期临床试验

Ⅱ期临床试验主要对新药的有效性、安全性进行初步评价，确定给药剂量。一般采用严格的随机双盲对照试验，以平行对照为主。通常应该与标准疗法进行比较，也可以使用安慰剂。需注意诊断标准、疗效标准的科学性、权威性和统一性。要根据试验目的选择恰当的观测指标，包括诊断指标、疗效指标、安全性指标。选择指标时，应注意其客观性、可靠性、灵敏度、特异性、相关性和可操作性。参照临床前试验和Ⅰ期临床试验的实际情况制定药物的剂量研究方案。应有符合伦理学要求的试验中止标准和退出标准。对不良事件及不良反应的观测、判断和及时处理都应作具体规定。应有严格的观测、记录及数据管理制度。试验结束后，对数据进行统计分析，对药物的安全性、有效性、使用剂量等作出初步评价和结论。Ⅱ期临床试验的试验组和对照组的例数都不得低于100例。

（三）Ⅲ期临床试验

Ⅲ期临床试验为扩大的、多中心随机对照临床试验，旨在进一步验证药品的有效性和安全性，评价利弊，最终为药物注册申请的审查提供充分的依据。可根据试验目的调整受试者的入选标准，适当扩大特殊受试人群，进一步考察不同对象所需剂量及其依从性。Ⅲ期临床试验的试验组例数一般不低于300例，对照组与治疗组的比例不低于1∶3，具体例数应符合统计学要求。

（四）Ⅳ期临床试验

Ⅳ期临床试验是在新药上市后研究，旨在评估药物在更大范围、长期的实际应用中的疗效以及监测不良反应，特别是罕发、严重的不良反应事件。此外，还应进一步考察对患者经济水平与生存质量的远期影响。Ⅳ期临床试验应在多家医院进行，观察例数通常不少于2000例。

临床试验的样本量除满足上述最低标准外，还需要同时满足统计学要求。由于新药临床试验是应用尚在研究中的新药，在人体进行的安全、有效性评价，故研究者对受试者的安全负有重要责任。因此需要在国家批准认证的药物临床试验机构（GCP）进行。

第二节　治疗性研究的设计内容及方法

临床治疗性研究主要包括以下内容：

一、确立研究目的及研究题目

一项高质量的临床治疗性试验，必须要有明确的研究目的。基于研究目的明确研究方案、设立指标、方法及统计分析手段。

在上一节已涉及临床治疗性试验的选题问题，强调选题应具有一定的创新性。题目可以来源于基础研究的提示，也可以是动物试验结果在人体的进一步验证，更多的是来源于临床医生

笔记

的实际观察和总结以及来自对人群流行病学的观察和研究。无论来自于哪一方面，其创新性、实用性和可操作性是必不可少的，而且一定要满足伦理学要求。

在立题过程中，需要关注的是采用什么样的方法、策略和技术路线来解决所发现的问题。不要匆忙立题，可以反复酝酿思考，并且最好让合作者再审查、评价，进一步把握研究问题的内容和方向，以保证整个研究能具有连续性。

二、选择研究方案

治疗性研究中多选择试验性研究方案，包括随机对照试验、非随机对照试验、历史性对照试验、自身前后对照试验、交叉对照试验及序贯试验等。其次还可以采用非试验性的研究方法，如队列研究等。

（一）试验性研究方案

治疗性研究可以采用的试验方案种类主要包括随机对照试验（randomized controlled trial，RCT）、单病例随机对照试验（N of 1 RCTs）、交叉对照试验（cross-over design，COD）、前 - 后对照试验（before-after study）、非随机对照试验（non- randomized controlled trial，NRCT）、历史性对照试验（historical control trial，HCT）、序贯试验（sequential trial）等。有关这些试验方案的定义、特点及应用条件在本教材第四章中已有详细介绍，在此不再作赘述。

（二）非试验性设计方案

此类代表性方案有队列研究（cohort study）和病例对照研究（case-control study），由于属于非试验性研究，其证据强度要弱于试验性的研究。此类内容在第四章中也有介绍，在此也不作叙述。

（三）多中心临床试验（multi-center clinical trial）

指由一个或几个单位的主要研究者总负责，多个单位的研究者合作，按同一方案进行的临床试验；各中心同期开始与结束试验，具有试验时间短、研究范围广、样本代表性好、结论外推性强等特点。但由于参与研究的单位和人员较多，对质量控制和标准化有更高的要求，同时，完成试验也需要较多研究经费。详见第四章。

（四）社区干预试验

是以社区自然人群作为研究对象，在现场环境下进行的干预研究，接受干预的基本单位是整个社区，或某个人群的各个亚人群，这种试验常用于对某种预防措施或方法（疫苗、药物等）的效果评价。

三、选择研究对象

研究对象首先应符合公认的临床及有关金标准诊断，同时根据研究目的，进一步拟定纳入标准及排除标准，以保证入选对象具有较高的同质性，有利于减少偏倚的发生，获得较为准确的研究结果。在设置排除标准时，应特别列出不宜使用该药的人群，如心、肺、肝、肾功能不全者和小儿、孕妇、哺乳期妇女等；同时对该类药物过敏和其他不宜参加这项研究者，如依从性差、参与了其他药物临床试验的受试者也应排除。但要注意：纳入标准的制定不宜过严，排除标准也不宜过多，否则就可能影响研究结果的外推性及适用性。

除上述要求外，还应确定病例的来源，包括是哪一级医院，是门诊病人还是住院病人。此外，根据医学伦理学的原则，凡参加临床试验者，都要签署知情同意书。

例如：某一新型抗生素治疗急性细菌性感染的疗效研究中，设置的纳入标准包括：年龄 18~65 岁的住院或门诊患者（慢性支气管炎急性发作部分患者年龄可放宽至 70 岁）、性别不限、经临床和实验室检查确诊为细菌感染、试验前未用过其他抗生素治疗或经其他抗生素治疗无效而细菌学检查为阳性并单独用此药治疗者、签署了知情同意书的受试者。并同时规定了排除标准：如

笔记

对抗生素有过敏史者；严重心、肝、肾功能不全者；有精神、神经系统疾患以及晚期肿瘤患者；妊娠、哺乳期妇女；依从性差或病情严重，不能完成试验者等。

四、研究对象的随机化分配及设盲

符合诊断标准、纳入/排除标准的合格受试者一旦入选药物临床试验，需进一步实施随机化分组，可选择的随机化方法，包括简单随机、区组随机、分层随机等，其中较为常用的是通过计算机产生随机序列或随机数字表产生随机数字。

随机分配方案的隐藏（allocation concealment），顾名思义就是对分配方案设盲，使研究参与者无法知晓某个患者所接受的是何种干预方案。随机分配隐藏的最佳方法是中心随机化，如将产生的随机分配序列号用信封密封。

此外，为避免临床试验中的测量性偏倚，也需要设立盲法。根据设盲的对象和范围，可分单盲临床试验（single blind method），双盲临床试验（double blind method）和三盲临床试验（triple blind method），其中以双盲临床试验最为常用。

五、估计样本量

通常在计算样本量时要考虑下述因素：

（一）最小临床效应量的设定

样本量估算首先应考虑疗效显著性差异水平的假设，确定临床最小效应量，即试验组与对照组的疗效差异多大才有实际意义或临床价值。计数资料的效应量表达采用试验组和对照组疗效率的差值，计量资料则为试验组和对照组的均值差值。

（二）Ⅰ型（α）错误和Ⅱ型（β）错误的水平

1. Ⅰ型（α）错误　即试验的假阳性错误，通常限定不超过0.05（5%）。该型误差越小，所需要的样本量越大。

2. Ⅱ型（β）错误　即试验的假阴性错误，通常限定为0.1，不宜超过0.2。$1-\beta$为检验效能（power），又称把握度。$1-\beta$越大，要求的样本量也越大。

（三）总体率π或标准差σ

样本含量估计时，对于计数资料，还需要确定总体率；对于计量资料，则需要估计标准差σ。

1. 两组率的比较　根据上面的这些基本设计参数，应用式（10-1）计算出各组所需的试验样本量。

$$n=\frac{\pi_1(100-\pi_1)+\pi_2(100-\pi_2)}{(\pi_2-\pi_1)^2}\times f(\alpha,\beta) \qquad 式(10\text{-}1)$$

其中n为一个组的样本量；π_1，π_2为试验组和对照组的发生率；$f(\alpha,\beta)$为限定假阳性和假阴性水平时相应的参数值，可以由表10-1查出。

表10-1　常用$f(\alpha,\beta)$数值表

α	β			
	0.05	0.10	0.20	0.50
0.10	10.8	8.6	6.2	2.7
0.05	13.0	10.5	7.9	3.8
0.02	15.8	13.0	10.0	5.4
0.01	17.8	14.9	11.7	6.6

例如：用两种药物对糖尿病患者进行康复治疗，经初步观察发现，甲药的有效率为70%，乙药的有效率为90%，现要进一步试验，设$\alpha=0.05$，$\beta=0.1$，问每组至少需要观察多少病例？

笔记

已知 $\pi_1=70\%$，$\pi_2=90\%$，$\alpha=0.05$，$\beta=0.10$，查表 10-1 得，$f(0.05,0.1)=10.5$，代入上述公式得：

$$n=\frac{70\times(100-70)+90\times(100-90)}{(90-70)^2}\times 10.5=79$$，即每组需观察 79 个病例。

2. 两组均数的比较 在进行两个样本均数比较的研究中，其样本量估计可采用式(10-2)。

$$n_1=n_2=2\left[\frac{(t_\alpha+t_\beta)S}{\delta}\right]^2$$ 式(10-2)

此式中的 n_1、n_2 分别为 1 组、2 组的样本量，一般是相等的，S 为两总体标准差 σ 的估计值，一般假设其相等或取合并方差的平方根，δ 为两均数的差值，即最小效应量的估计，t_α 和 t_β 分别为检验水准 α 和第Ⅱ型错误的概率 β 相对应的 t 值，可以通过查 t 值表得到。

例如：观察两种药物治疗肌痉挛的疗效，其中 B 药使肌痉挛分数平均减少 2.16，L 药使肌痉挛分数平均减少 1.66，设两种药物疗效的标准差相等，均为 0.7 分，要求 $\alpha=0.05$，$\beta=0.1$，若要得出两组处理差别有显著性结论，需要多少研究对象？

已知：$\delta=2.16-1.66=0.5$，$S=0.7$，双侧 $\alpha=0.05$，$\beta=0.1$，查 t 值表得：$t_{0.05/2,\infty}=1.96$，$t_{0.1,\infty}=1.28$，代入公式得：

$$n_1=n_2=2\left[\frac{(1.96+1.28)\times 0.7}{0.5}\right]^2=41.2=42(\text{例})$$

故认为两组各需 42 例患者，共计 84 例。

在实际工作中，可以采用下述公式进行计算，较为方便。

$$n=2\hat{\sigma}^2\times f(\alpha,\beta)/(\mu_2'-\mu_1')^2$$ 式(10-3)

n 为每组所需的例数，μ_2'，μ_1' 分别为两组的预期均数，$\hat{\sigma}^2$ 为两组的合并标准差或对照组的标准差，$f(\alpha,\beta)$ 可由表 10-1 中查出。用此公式计算结果相同。

考虑试验中出现的病例失访和退出情况，故常在计算量的基础上增加 10%。

此外，在实际工作中还可以根据不同研究方案的特点调整样本量，如非劣效性试验和等效性试验，所需样本量应适当增加。新药审批办法中虽已明确规定了新药临床试验的病例数要求，但也应同时满足统计学要求，需要估计样本量。

六、选择试验药物或措施

试验药物或措施首先要保证其有效性和安全性，并同时具有一定的创新性和经济性。对照组药物，无论是阳性对照药物还是安慰剂，在外观、色泽等方面应与试验组药物相似，同时服用方法和疗程也要一致，否则会影响结果的真实性。

七、选择试验疗效和安全性测试指标

临床试验中应采用某种测量方法和指标来度量药物或干预措施所呈现的治疗效应，包括疗效及药物不良反应等，所设置的测试指标应满足以下要求：

1. 灵敏度要好 对于治疗引起的客观反应，测试指标要能敏感地发现并能度量。例如：采用胆碱酯酶抑制剂抗痴呆治疗，如果应用临床痴呆分级表，其敏感性要比阿尔茨海默评估量表认知副表要差得多，如用前者测量治疗反应，必然大大增加假阴性率，所以测试指标和方法的灵敏度越高越好。

2. 特异度要强 测试方法和指标除了敏感之外还要特异。例如，溶栓疗法治疗急性心肌梗死，在治疗前后采用冠脉造影，比较分析冠脉狭窄和闭塞改善的程度，若以此作为疗效的测量指标，则特异性强，有助于疗效的判定。

笔记

3. 针对性要佳　测试指标的选择应充分考虑试验的终点目标。如治疗的终点目标是降低病死率和非致死事件发生，则测试指标定为病死率、生存率以及非致死事件发生率（如冠心病发生心肌梗死、心衰等）。若验证治疗措施本身的有效性，则采用临床公认的有效或无效的判断标准，精选有关临床及实验室的定量及定性指标予以测量。如用血压计测量降压效果，用血糖评价糖尿病降糖效果等。

4. 指标数量要适当　指标设置要少而精，测试指标越多，假阳性的发生概率就越大，有时甚至会左右研究结论。

5. 远期效应指标要优先　对于某些慢性病的疗效评价，除测试和评价近期效应外，还要追踪观察远期效果，这样更有助于临床决策。例如，高血压患者的治疗不仅要观察用药后血压的控制水平，而且更要观察高血压所致的心、脑、肾不良事件的发生情况。

6. 经济可行性　测试方法和指标除了要考虑敏感性和特异性外，还要注意这些指标的经济可行性。如评价输卵管堵塞病人的复通效果，可选用输卵管子宫碘油造影或腹腔镜下通亚甲蓝等两种方法，这些方法的敏感度和特异度均较高，但前者更为经济且操作相对简便，对病人的创伤小，故应优先考虑。

7. 指标设置要全面　在药物临床试验中，除了有效性指标外，安全性指标同样重要，主要包括一些症状、体征、实验室检查指标。选用时应根据试验目的、内容而定。

八、确定干预实施方法及随访观察期

详细说明试验干预方法和实施要求，建立试验药物包装、分发、转运、供应档案。设立盲法时，则有保证盲法的具体措施，以及当患者病情恶化或突发不良事件，有紧急破盲的操作规范；同时还要建立避免沾染和干扰以及保证依从性的制度等。

临床试验中同样需要确立一个明确的随访观察期，观察期要适当、不宜过长或过短。过长会造成不必要的浪费，过短则可能会无法观察到一些远期效果。观察期的确定应基于研究目的、前期基础研究结果以及临床达到治疗最佳水平所需时间等。如骨质疏松的防治性研究应考虑到骨代谢的周期较长，少于1年很难得出结果。

九、整理分析试验结果

与其他研究一样，临床试验也需在原始资料完整、准确的基础上，按研究目的、试验设计方案，以及资料类型选择正确的统计方法处理资料。内容包括：详细列出主要和次要指标的分析方法、亚组分析方法；详细比较进入试验组和对照组病人的基线特征，以及失访、退出和脱落病例的情况；在分析中如何处理偏倚和缺失数据，如何解释结果的意义；如实报告试验结果的有效性、安全性等。

（一）资料的整理

对获得的原始记录包括病历、观察表、临床化验及各种功能检查结果要进行核查，然后通过计算机建立数据库、双输双录所有的资料。对于可疑的数据可以下达数据质疑表，经修正后再次修订数据库。当所有的修订完成后，锁定数据库。

（二）统计方法的选择

1. 根据资料的性质采用相应的统计学方法

（1）治疗性研究资料中最常见的是计数资料与计量资料和等级资料。计数资料的描述采用率和比，如有效率、治愈率、病死率等，假设检验方法为卡方检验（χ^2 test）；计量资料的描述常用均数 ± 标准差，假设检验常用 t 检验（小样本），u 检验（大样本）及 F 检验（多因素方差分析）、非参数检验等；等级资料是将某一指标划分为若干等级，常用 Ridit 分析及非参数检验等。

（2）如治疗性研究本身有两组以上组别比较，必须先作多组间的显著性检验，差异有统计学

笔记

意义时，才能进一步作两两比较。

（3）治疗性研究设计可以是配对设计，或完全随机设计。两种设计的原理不同，分析处理的方法也不同，二者不可混淆。

2. 根据分析目的选用统计方法

（1）治疗效果的多因素分析：任何治疗效果的产生，除了治疗措施本身的效力之外，还与患者人文特征以及病理生理状态、疾病特点等有关。例如年龄、营养状态、病情、药量、疗程、并发症、合并症等，均会影响治疗反应。为此，应在单因素分析的基础上，选择具有显著意义的有关变量作多因素分析，进一步评价疗效。

（2）单侧检验或双侧检验：如肯定试验药（或措施）疗效优于对照药（或措施），则用单侧检验法；若不能肯定，则采用双侧检验法。如果对照组用阳性对照药物，则需要进行非劣效性检验，此时用单侧检验即可。

（3）对不完整资料的处理与分析：对于在研究中出现退出、失访的情况，除了采用按方案分析（per-protocol analysis，PP）之外，还可以采用意向治疗分析（intention to treat analysis，ITT）。ITT分析是将所有纳入随机分配的病人，不管最终是否接受到分配的治疗，在最后资料分析中都应按随机分配方案统计，以保证结论更真实可靠。

第三节　治疗性研究质量的常见影响因素及其处理方法

治疗性研究中存在很多影响结果真实性和可靠性的因素，如常见的偏倚和机遇因素等，若不加以识别和控制，临床研究结果将会失真。

一、机　遇

机遇（chance）即随机误差或抽样误差。机遇因素在治疗性研究中不可能消除，只能在研究设计阶段，通过限制Ⅰ型错误率和Ⅱ型错误率，使机遇因素的影响控制在容许范围之内。

二、偏　倚

除常见的选择偏倚（selection bias）、测量偏倚（measurement bias）和混杂偏倚（confounding bias）外，还存在一些临床研究特有偏倚。

（一）选择偏倚

该偏倚主要产生在设计阶段，入选研究对象与未入选者在某些特征上存在系统差别，同时除研究因素以外的其他因素组间分布也不均衡，导致研究结果出现偏差。选择偏倚可以通过随机、设立对照、严格纳入/排除标准等方法加以控制。

（二）测量偏倚

测量性偏倚指在资料的观察、测量及收集过程中，由于受到人为因素的影响导致结果失真的一类偏倚。通过实施盲法、标准化测量、提高应答率等方法可以防止这类偏倚的发生。

（三）混杂偏倚

混杂偏倚是指由于混杂因素的存在，掩盖或夸大了研究因素与疾病间的关联度。可以通过限制、配比、随机、标准化、分层及多因素分析方法加以控制。

（四）特殊偏倚

在临床治疗性研究中还存在一些特殊偏倚，在本质上属于测量性偏倚，主要包括以下几种：

1. 霍桑效应（Hawthorne effect）　霍桑效应是指在研究过程中，研究者对自己感兴趣的研究对象较对照者往往更为关照和仔细；而被关照的患者对研究人员又极可能报以过分的热情，更多地向医生报告好的结果。这种人为地夸大客观效果的现象，称为"霍桑效应"。控制霍桑效应

笔记

的最好方法是严格实施盲法。

2. 干扰(co-intervention) 干扰是指试验组或对照组的对象额外地接受了类似试验药物的某种有效制剂，从而人为地夸大了疗效的假象。如果是试验组的对象接受了“干扰”药物，则会造成疗效提高，从而引起试验组与对照组疗效差异的增大；反之，如果是对照组的对象接受了“干扰”药物，则可引起对照组疗效率增高而使组间疗效差异减小。最好的控制方法就是实施规范化试验及改善依从性，尽可能减少其他药物的使用。

3. 沾染(contamination) 沾染是指对照组患者额外地接受了试验组的药物，人为地夸大了对照组疗效的现象。在试验设计时应该加以限制并在试验过程中加强质量控制，改善依从性。

4. 向均数回归现象(regression to the mean) 有些测试指标如血压或某些生化指标在初试时有些患者可以在异常水平，然而，在未干预或无效治疗的条件下复试，可能有些回复到正常水平。这种现象或许属生理性波动，而非干预的结果，但可造成治疗有效的假象。克服的办法是可以采取对同一个体的有关测试指标在相同条件下，不同时间内多次测定，取均值以排除其干扰。

5. 失访 当大量病人失访时，研究也可以产生与结局测量有关的偏倚。因为失访病人如果有不同的事件发生率，仅仅依靠随访到的病人的数据资料将会得到与真实情况不同的结果。克服的方法是提高随访率或者采用ITT分析。

6. 依从性(compliance) 是指研究对象按照研究设计要求执行医嘱的客观反应的程度。全面认真地执行医嘱，按规定的药物剂量和疗程接受治疗，称为依从性好；反之则为依从性不好或不依从(non-compliance)。依从性差的原因包括：简单的遗忘、误解药物使用方法、不能耐受药物的副作用、讨厌服药或费用不足等。治疗时间长(如几个月)或治疗方案复杂对依从性也有较大的影响。

解决依从性问题最主要的方法是使病人充分理解试验目的、要求及参加这项试验的意义，使病人在理解的基础上给予合作。此外，还必须同时加强试验工作的管理，从客观上减少不依从的可能性。如建立检查制度，在复诊时计算患者的依从性，见式(10-4)，如服药量≥80%，则依从性佳，<80% 为依从性不佳。为了保证研究质量，不依从率应力争控制在10% 范围内。同时还应建立药物、血、尿浓度的检测方法，必要时可进行体液内药物浓度测定来确保病人的依从性。另外，选用高危人群作研究对象对提高依从性也很有意义。

$$\text{依从性}=\frac{\text{实际用药量(应服药量}-\text{复诊时剩余药量)}}{\text{总设计用药量(应服药量)}}\times 100\% \qquad \text{式(10-4)}$$

患者的依从性程度对研究的质量会有很大的影响。依从性越高，治疗组和对照组间的差异越能反映治疗的真正效果。而依从性降低就不可能获得预期效果，使本来应该有的治疗效果消失或降低。所以在治疗性试验中，维持与改善患者的依从性是十分重要的。

第四节　治疗性研究的评价原则

治疗性研究结果只有通过科学地评价才能区别措施的利弊以及决定能否用于临床实践。无论是单个原始研究评价还是系统综述或 meta 分析等二次研究评价，都可以概括为内在真实性(internal validity)、重要性(importance)、适用性(applicability)三个方面。

一、单个原始研究评价

(一) 真实性评价

1. 治疗性研究是否为真正的随机对照试验 随机对照试验的设计要求最为严格，结果的真实性优于其他设计方案。这是因为设计严谨的RCT可有效地控制已知或未知的偏倚的干扰，

笔记

确保研究结果真实可靠，故为治疗性研究的金方案。在评价 RCT 时要考虑以下问题：①是否采用了真正的随机方法，是否交代了具体的随机方法，是否实施随机分配隐匿，是随机还是随意。对于分层随机对照试验，要注意分层因素的数量、试验组与对照组的样本量及分层后各亚组的病例分布是否一致。②是否采用了盲法，要注意是否交代了具体的"盲"法，实施的是单盲、双盲或三盲，阐明试验过程是否按盲法操作等。③组间基线状态的可比性如何，要注意是否描述了组间基线的状态并作了比较，如果存在组间基线状态的不一致性，则要注意是否作了分层比较或校正。④伴随的辅助治疗是否对结果有影响，在某些较为复杂疾病的治疗性研究中，有时是在"基础治疗"的基础上进行试验干预的，或许在试验中某些病例出现新征候需要辅以相关的临时用药。因此，组间的辅助或基础治疗，应注意差异，此外，还要注意考证有无"干扰"及"沾染"的影响。

如果结果是来源于非 RCT 研究，则应对文献进行分析与评价，以获取优质的证据。同时还要考虑其他的注意问题。

2. 所有纳入的研究对象是否随访完整？研究对象随访时间是否足够 任何观察病例的丢失，都会直接影响最后的结果和证据的真实性。例如，疗效差的患者退出，会导致治疗效果高估；若个别患者因药物或者干预措施的副作用从治疗组中退出，可能会低估药物的危害性。理想的情况是所有纳入的研究对象在研究过程中都没有失访，但这一点在实际临床研究中是很难保证的。

失访率一般应控制在 10% 以内，若超过 20%，研究质量会受到很大影响，结果也许会不真实。对于一般失访率影响的估计，常用的方法是将试验组全部失访的病例，均计算在本组无效病例中，对照组丢失的病例全部计入有效病例内。如果仍然与原有效结论一致，则可以接受原来的结果。

同时，应该确保随访期足够长，以保证获得重要的临床效应结果。随访的具体时间取决于目标疾病的病程特点，通常临床观察的疗程至少数月，有的甚至需要 1 年以上方能充分显示防治措施的重要效果。

3. 是否对随机分组的所有研究对象进行了分析（是否采用意向治疗分析） 被随机分配入组的病例可因为各种原因出现失访。例如，因副作用发生出现中途停药者；患者依从性差而未认真按医嘱服药者以及发生沾染或者干扰者等。如果这部分研究对象不被纳入结果分析，必然影响原来的随机化原则和基线的可比性，最终影响结果的真实性。目前广泛采用意向性治疗分析进行资料分析，即按最初随机分配入组的全部病例，无论其是否接受或未接受确切的治疗药物，都纳入进行最后分析。

4. 是否对研究对象、医生和研究人员实施了盲法 在治疗性资料的收集过程中往往也会产生较大的测量性偏倚。如受试研究对象知道自己接受的是治疗措施、还是对照措施；研究者或结果测量者知道研究对象的分组情况，往往会夸大研究药物的效应等。实施盲法的目的就是为了减少测量性偏倚以维护观察结果的真实性。盲法可以是单盲、双盲或三盲。

对于无法实施盲法的试验（如外科手术），可以请另外的医生采用客观指标评价治疗效果。若使用临床记录，则应去除所有可能涉及破盲的治疗措施的信息，以保证真正的采用了盲法。

5. 除试验方案不同外，各组患者接受的其他治疗方法是否相同 如果研究对象除了接受规定的治疗方案外，还有意或无意采用了其他的具有类似作用的干预措施，必然影响结果的真实性。沾染和干扰即是常见的两种情况。前者是指对照组的患者接受了试验组的防治措施，使得试验组和对照组间的疗效差异减小；后者是指试验组或对照组接受了类似试验措施的其他处理，人为扩大或减小组间疗效的真实差异。

如果一项随机对照试验不符合上述五条标准的一条或多条，则需要考虑其研究结果的真实性。

笔记

（二）重要性评价

对于治疗性证据的重要性评价应注重两个方面，即正面的有效性和负面的不良反应，只有疗效佳、负效小者方有临床价值。同时还要分析其精确度。

1. 治疗性证据的效应强度大小

（1）疗效强度：疗效的强度通常用率表示，即有效率、治愈率、病死率、病残率，即使有些计量的疗效指标，亦多转换为"有效率"等。

当对RCT各组的有效率进行比较分析时，可以采用统计学方法判定各组间差异的显著性，然而，仅仅采用这些"率"对临床量化其重要程度的表达是不足的。近些年来，随着临床流行病学与循证医学的发展，一些能够更好反映临床意义的试验结果指标被广泛应用，主要有以下几种：

1）相对危险降低率（relative risk reduction, *RRR*）

$$RRR=\frac{CER-EER}{CER} \quad 式(10\text{-}5)$$

注：*CER* = control event rate（对照组事件率）；*EER* = experiment event rate（试验组事件率）。

2）绝对危险降低率（absolute risk reduction, *ARR*）

$$ARR=CER-EER \quad 式(10\text{-}6)$$

3）需治疗多少病例数才获得一例好结果（number need to treat, *NNT*）

$$NNT=\frac{1}{ARR} \quad 式(10\text{-}7)$$

例如：一个应用他汀治疗预防脑卒中的RCT中，治疗5年后他汀组的脑卒中发病率为4.3%（*EER*），对照组的脑卒中发病率为5.7%（*CER*），根据以上公式，计算进一步量化的指标为：

$$RRR=\frac{5.7\%-4.3\%}{5.7\%}=25\%, ARR=5.7\%-4.3\%=1.4\%, NNT=\frac{1}{1.4\%}=72$$

（2）负效值的强度：通常某种新药的临床治疗性试验，特别是与安慰剂比较研究时，新药的不良反应往往较对照组为显著，也许还可能发生较重的药物不良反应（ADR）。因此，在分析和评价治疗性证据时，要注意不良反应的各组发生率、种类及其强度，如同分析评价正面疗效一样，进一步测试：

1）相对危险增加率（relative risk increase, *RRI*）

$$RRI=\frac{EER-CER}{EER} \quad 式(10\text{-}8)$$

2）绝对危险增加率（absolute risk increase, *ARI*）：

$$ARI=EER-CER \quad 式(10\text{-}9)$$

3）需治多少病例才发生一例不良反应（the number needed to harm, *NNH*）

$$NNH=\frac{1}{ARI} \quad 式(10\text{-}10)$$

例如：上述他汀RCT，试验组不良反应率0.05%，对照组0.03%，则该研究之不良反应的：

$$RRI=\frac{0.05\%-0.03\%}{0.05\%}=40\%, ARI=0.05\%-0.03\%=0.02\%, NNH=\frac{1}{0.02\%}=5000$$

由此可以看出，他汀治疗的不良反应发生率很低，相对而言是较为安全的。

2. 治疗性结果的精确度 研究中采用的各相关指标，仅表示着它们效应强度的大小，如有效率80%以及*ARR*、*RRR*、*NNT*的值等。但效应量的精度往往会因不同样本量大小而有所差异。因此，应进一步评估效应的精确度（precise）。

笔记

通常采用95%的可信区间(95% *CI*),其波动范围代表结果的精确程度,95% *CI*值的范围越窄则越精确。

(三)适用性评价

上述证据的真实性与重要性经分析和评价获得肯定结论之后,紧接着要考虑这种有价值的证据是否可被应用于临床实践,需要结合自己病人的实际情况,评价它的适用性。通常考虑以下几条规则:

1. 被评价的证据是否与病人情况不符而不能应用

(1)整体证据:任何治疗性研究所产生的有价值的证据,应该审查疾病的诊断标准是否可靠,证据中的研究对象之纳入标准是否与拟引证的病人相符,其生理功能与病理学的依据、病情特点、年龄、性别以及社会经济状况是否存在显著差异等,假若以上特点一致或大体一致,则该治疗性证据就认为适用,否则就不可取。

(2)亚组证据:有的证据在总体上可能缺乏适用性,但其亚组分析的疗效可能有显著意义,如果患者的病情与某亚组患者的病情相似,那么,这个亚组的治疗证据就有适用的价值。

对于有着显著适用的亚组分析证据,如能符合下述条件者,则有应用价值:

1)确有生物学和临床依据者;

2)确有统计学意义和临床价值者;

3)亚组分析在研究设计阶段就已拟定,而不是在试验结束后有意为之;

4)仅属该研究中有限几个亚组分析的证据之一者(注意亚组不能设置过多);

5)该证据在另外的研究中可被证实者。

2. 拟采用的治疗证据是否可以在本单位医疗条件下采用 对于拟采用的有效治疗措施,需要在具有一定医疗水平、医疗条件的医院才能被采用,如医生的技术水平,医院的管理机制及设备条件、病人的意愿以及经济的承受能力等。如像冠心病的介入治疗、风湿性心瓣膜病的换瓣治疗术等,即使这类治疗证明对病人有利并具颇佳效果,但不具备上述条件,也是不可行的。

3. 从该治疗性证据中,估计患者的利、弊 采用某一可行性好的最佳治疗措施一定是要保证利大于弊,且要有量化指标为依据。最直接的是用*NNT*(益处)及*NNH*(害处)进行评价。如果这两种指标不全,则有两种办法帮助解决。

一是确定患者预期事件发生率(patient's expected event rate, *PEER*),意思是如果患者不予治疗,其最终结局事件的发生率。这可以用治疗试验中的安慰剂对照组的事件发生率(*CER*)表示;如无*CER*证据,也可根据临床积累的未治或缺乏特效治疗观察结果值作为*PEER*参考值,如像急性心肌梗死患者*PEER*约15%;另外也可以亚组资料分层分析之相应的*CER*作为*PEER*等。

当获得*PEER*、*RRR*及*RRI*等指标值后,用下列公式则可推算:

$$NNT=\frac{1}{PEER \times RRR} \qquad 式(10\text{-}11)$$

$$NNH=\frac{1}{PEER \times RRI} \qquad 式(10\text{-}12)$$

二是应用列线图,用已知的*PEER*(或称不治疗的绝对危险率)与*RRR*数据,依据此两值直线连线的延伸,与*NNT*线上交叉点,即为该*NNT*。(图10-1)十分方便。

4. 考虑患者对于治疗措施的价值取向与期望 在循证医学治疗实践中,一定要尊重患者对治疗的价值取向,即愿意接受或者不愿意接受,或愿意接受哪一种备选方案(药物或有关治疗措施),而且要了解患者对治疗结局的有关期望。

笔记

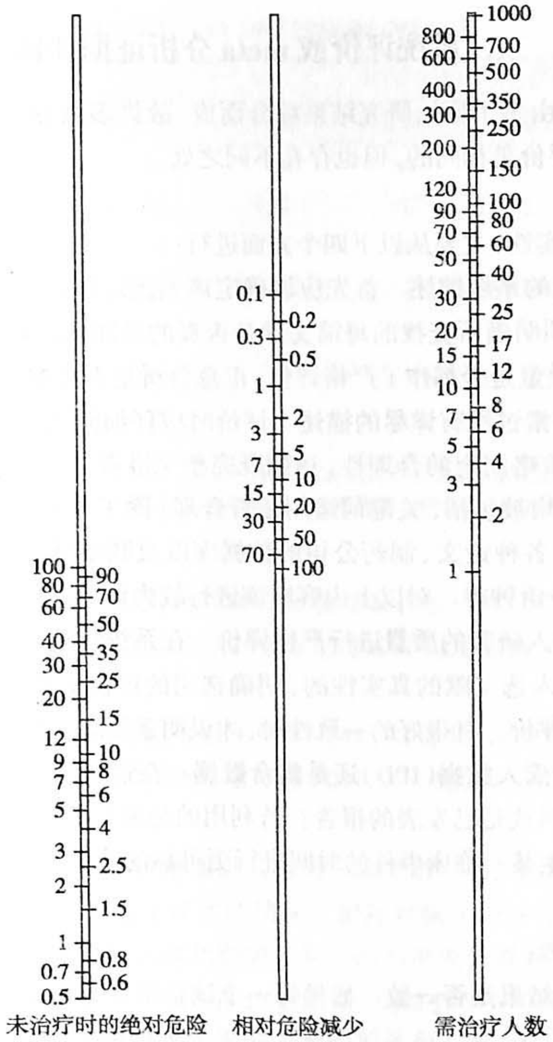


图 10-1　估计 NNT 的列线图

首先，要保证不良反应最小化；同时还要注重疗效。因此，估计与掌握治疗的利弊比就十分重要了。通常应用治疗措施（或药物）的 *NNT* 与 *NNH* 计算其利弊比（likelihood of being helped vs harmed，*LHH*）。

$$LHH=\left(\frac{1}{NNT}\right)\Big/\frac{1}{NNH} \qquad \text{式(10-13)}$$

例如：有关他汀药物治疗预防脑卒中的 *NNT* 为 72，*NNH* 为 5000，则

$LHH=\left(\frac{1}{72}\right)\Big/\frac{1}{5000}=70$，这意味着有利面是不利面的 70 倍，是为安全和有效的治疗证据，显然 *LHH* 是越高越佳。

如果有几种备选药物同时存在，且它们的疗效与不良反应相似或有差异，但没有显著意义。对这些备选治疗措施（或药物），则应优先选择其成本（价格）低廉且疗效也好和安全的药物，在保证安全有效的基础上，尽可能地降低医疗成本。同时，对于任何治疗措施，一定要给患者尽可能的清楚解释，以保持良好的依从性。并在治疗过程中务必要认真观察治疗反应，关心与爱护患者，这有利于增进互信和睦关系，避免产生不必要的误解或纠纷。

笔记

二、系统评价或meta分析证据评价

系统评价（综述）是由多个原始研究结果综合而成，故许多内容（尤其是评价重要性与适用性的部分）与原始研究评价是相同的，但也存在不同之处。

（一）真实性评价

评价系统综述的真实性，主要从以下四个方面进行：

1. 是否是随机试验的系统综述　首先应该确定该系统综述是针对随机试验还是非随机化试验。如果是前者，必须明确所查找的每篇文献是否真的是随机试验，其中是否包含了非随机试验；对每篇合格文献质量是否都作了严格评价；汇总分析是否考虑纳入研究结果间的异质性。

2. 是否对文献的检索过程有详尽的描述　评价时应仔细阅读总结报告中与检索有关的方法学部分，特别是检索策略制定的合理性，达到既完整又准确。包括检索范围是否广泛，主要的医学文献数据库是否均被包括，关键词运用是否合理；除了计算机检索外，是否采用包括手工检索期刊、会议记录、各种论文、制药公司的数据库以及联系已发表文章的相关作者等多种手段；是否只局限于单一语种等。对以上内容应该进行较为详尽的描述。

3. 是否已对每个纳入研究的质量进行严格评价　在系统综述的方法学部分还应当说明调查者是如何评价每一篇入选文献的真实性的，明确选用的评价标准，文献纳入的方法，即是否由两人或多人独立进行评价。有很好的一致性时，才说明系统综述结果更加可信。

4. 分析时采用原始病人数据（IPD）还是集合数据　在进行分析时采用的数据是来自原始的病人资料还是表格资料或是已发表的报告？若利用的是原始数据，则综述结论更可信。使用原始资料还可对病人发生某一临床事件的时间进行更可靠的分析。

（二）重要性评价

重要性评价主要考虑：

1. 不同原始研究的结果是否一致　如果每一个试验结果都显示了相似的效果或者至少方向相同，即有较好的同质性，那么该系统综述的结果令人信服。但在实际研究中，达到完全的同质是不可能的，表现为部分研究的结论可能有效，部分为无效，甚至有害。此种情况下更需进一步分析，找出异质性的来源。

2. 治疗效果有多大　有较多指标可以用于系统综述的效果量表达，如*NNT*、*RRR*、*ARR*、*RRI*、*ARI*、*NNH*等。目前已有许多系统综述用*NNT*来表示结果，因为*RRR*未考虑*CER*和*PEER*，*OR*和*RR*也因同一原因在临床实践中应用较少。现已有一些工具将*RR*、*OR*转换为*NNT*，部分转换结果可参见表10-2及表10-3。

表10-2　*OR*-*NNT*转换表（*OR*<1）

病人预期事件发生率（*PEER*）	*OR*值						
	0.9	0.8	0.7	0.6	0.5	0.4	0.3
0.05	209[a]	104	69	52	41	34	29[b]
0.10	110	54	36	27	21	18	15
0.20	61	30	20	14	11	10	8
0.30	46	22	14	10	8	7	5
0.40	40	19	12	9	7	6	4
0.50	38	18	11	8	6	5	4
0.70	44	20	13	9	6	5	4
0.90	101[c]	46	27	18	12	9	4[d]

注：(a)此处相对风险减少（*RRR*）为10%；(b)此处*RRR*为49%；(c)此处*RRR*为1%；(d)此处*RRR*为9%。引自John Geddes.personal communication，1999

笔记

表 10-3　*OR-NNT* 转换表（*OR*>1）

病人预期事件发生率（*PEER*）	*OR* 值 1.1	1.25	1.5	1.75	2	2.25	2.5
0.05	212	86	44	30	23	18	16
0.10	113	46	24	16	13	10	9
0.20	64	27	14	10	8	7	6
0.30	50	21	11	8	7	6	5
0.40	44	19	10	8	6	5	5
0.50	42	18	10	8	6	6	5
0.70	51	23	13	10	9	8	7
0.90	121	55	33	25	22	19	18

注：表中数字是在特定 *PEER* 水平时，*OR* 值对应的 *NNT*。此表适用于治疗引起的效应与副作用，引自 John Geddes. personal communication，1999

3. 证据效果的精确性如何　与对单项研究证据结果评价一样，仍需要采用 95% 可信区间来评价系统综述效果的精确性，以进一步估计效果的强度及其临床意义大小。

（三）适用性评价

由于系统综述中的各个研究可能来自不同的人群，其合并效应估计值是平均的疗效。因此，即使系统综述结果已经具备了真实性、重要性，但也不一定适用于具体的病人，仍需要进行适用性评价。

系统综述适用性评价原则与评价单项研究相同。即考虑：自己诊治的病人与研究中的病人是否存在很大的差别？治疗是否适于现实的情况？治疗对病人的潜在利益和损害有哪些？患者对治疗结局和提供的治疗方案的态度和期望如何？但与单项研究评价不同的是，系统综述会清楚地给出不同病人或不同情况下各亚组的信息结果，从而会更方便地应用于自己的病人。但在应用时必须满足亚组病人的应用条件。

不是所有的系统综述都是高质量的。在应用系统综述作为证据时，也应该采取严格评价的态度，切不可盲信。因为最佳的系统综述必须基于高质量的 RCTs，若 RCT 质量差、研制者专业水平不高、制作过程中偏倚多，其证据价值一定不高。

第五节　循证实践及案例分析

急性脑梗死（acute cerebral infarction）是各种诱发因素引起脑内动脉狭窄、闭塞，而造成局部急性脑血液循环障碍，引起缺血、缺氧导致局灶性脑组织的缺血性坏死、软化。有研究表明，40 岁以后，年龄每增长 5 岁，缺血性脑卒中发病率约增加一倍。急性脑梗死是神经系统的常见病，致死率及致残率均较高，对人类及社会危害巨大。目前，对于该病的主要治疗方法包括：溶栓、抗血小板、抗凝、脑保护剂、降纤治疗及手术治疗等。选择性血管内溶栓是治疗急性脑梗死首选方法之一，被认为能有效缓解局部脑组织缺血缺氧的情况，减少脑组织的梗死面积。但救治时间窗、患者基础情况不同会导致最佳的治疗措施不同，故针对临床上某一具体的急性脑梗死患者，医生常会遇到，如该患者最佳治疗措施是什么，患者是否应该使用溶栓治疗，应该选用哪种溶栓药，急性脑梗死患者如何处理血压等一系列问题。本节将以此实例进行循证实践。

一、疾病案例

患者女性，65 岁，右侧半身无力，说话不清，4 小时收入某院神经内科，有高血压病史 15 年。查体：神清，体温 36.5℃，脉搏 80 次 / 分钟，血压 165/95mmHg，发育正常，营养良好，头颅无畸

笔记

形，心肺腹部无异常，双瞳孔等大等圆，光反射正常，眼底阴性，脑神经阴性，有鼻唇沟变浅。检测：急诊脑部 CT 扫描没有出血和其他异常密度影；血生化及常规检查正常；全脑血管造影检查发现右侧大脑半球血流量明显减少。初步诊断为急性脑梗死。

二、提出问题

针对上述这位伴有高血压的急性脑梗死患者提出临床问题：能否使用溶栓药物治疗？是否应该降血压？其他疗法应该怎样选择？

三、检索证据

全面收集有关研究证据——查寻最新指南、系统综述、临床研究证据。英文以 acute ischemia stroke、thrombolysis、hypertension 为关键词组合分别检索 PubMed、Web of Science、Cochrane Library 等英文数据库；中文以中国学术期刊网（CNKI）全文数据库、重庆维普（VIP）中文科技期刊全文数据库、万方科技期刊全文数据库作为检索数据库，以急性脑梗死、急性脑缺血、溶栓及高血压作为关键词进行检索。制定检索策略时，注意检索词的变换使用，如 acute cerebral infarction、acute cerebral ischemia、thrombolysis therapy、blood pressure 等。检索结果：关于溶栓治疗急性脑梗死的国内外研究相对较多，其中检索到 3 篇最新相关指南、近百篇系统评价及 meta 分析，包括 8 篇在 Cochrane 图书馆注册的相关系统评价。其中，2003 年发布的美国和欧洲指南各一个，并分别于 2006 年和 2008 年更新；2007 年发布的中国脑血管病防治指南，并于 2010 年第一次修正。

四、评价证据

严格评价研究证据。根据具体的临床问题，寻找答案。指南及系统评价是在综合相应证据的基础上，具有较高的证据级别，其次是高质量的随机对照试验。针对本例中提出的问题，检索到了多个指南，对其给出的推荐予以重点考虑。

（一）能否采用溶栓治疗

有关指南和证据结论如下：

1. 指南均建议对缺血性脑卒中发病 3 小时内（Ⅰ级推荐，A 级证据）和 3~4.5 小时（Ⅰ级推荐，B 级证据）的患者，应根据适应证严格筛选患者，尽快静脉给予重组组织型纤溶酶原激活（rt-PA）溶栓治疗。
2. 在发病初期，尚无其他溶栓药物可以代替 rt-PA 的研究证据。
3. 其他溶栓药物只能在临床试验的情况下使用。

溶栓治疗的研究证据：Wardlaw 2009 年的 Cochrane 系统评价收集了全世界 26 个 RCT（包括 7152 例病人），结果为：溶栓治疗组致死性颅内出血较未溶栓组增加 5 倍，症状性颅内出血增加 3.5 倍，远期病死率约增高 1/3。但 3~6 个月后 6 小时内使用溶栓治疗者，死亡或残疾的危险降低 19%，3 小时内溶栓者似乎更为有效。结论是：①溶栓组早期死亡和颅内出血的风险增加，但这些风险可被存活者残疾率降低的效果所抵消。②不清楚哪些病人可能受益，哪些病人可能风险更大。③开始治疗的时间可能和预后有关。④ rt-PA 研究提示 3 小时内使用效果较好，6 小时内使用仅限于条件较好的机构，并且需非常谨慎的选择病例，大范围常规使用尚无充分证据。

（二）血压的处理问题

美国指南建议：①对急性缺血性脑卒中（脑梗死）患者的高血压处理应十分谨慎，首先应密切监测血压，处理焦虑、头痛、处理高颅内压。②经上述处理后血压仍持续升高 >220/120mmHg 时可用降压药，但须慎重进行，缓慢降压。③紧急情况下应使用非肠道给药。避免使用强作用的长效降压药物。④对有溶栓指征的患者，血压应降至≤185/110mmHg。欧洲指南建议：①脑梗

笔记

死血压＞200/120~220/120mmHg，脑出血患者血压持续＞180/105mmHg 时可用降压药。②有溶栓指征的患者，收缩压应降至 185mmHg 以下。③对有心衰、急性心肌梗死、急性肾衰使用溶栓或抗凝情况的患者，应立即降压。国内指南建议：①准备溶栓者，应使收缩压＜180mmHg、舒张压＜100mmHg。②缺血性脑卒中后 24 小时内血压升高的患者应谨慎处理。应先处理紧张焦虑、疼痛、恶心呕吐及颅内压增高等情况。血压持续升高，收缩压＞200mmHg 或舒张压≥110mmHg，或伴有严重心功能不全、主动脉夹层、高血压脑病，可予谨慎降压治疗，并严密观察血压变化，必要时可静脉使用短效药物（如拉贝洛尔、尼卡地平等），最好应用微量输液泵，避免血压降得过低。③有高血压病史且正在服用降压药者，如病情平稳，可于脑卒中 24 小时后开始恢复使用降压药物。④脑卒中后低血压的患者应积极寻找和处理原因，必要时可采用扩容升压措施。

（三）关于其他疗法

欧美的指南建议：①证据充分应广泛使用：卒中单元，阿司匹林 100~300mg/d（Ⅰ类证据，A 级建议）。②证据欠充分应限制使用：有高度深静脉血栓形成或肺栓塞风险的患者，建议应当考虑给予低剂量皮下肝素或低分子量肝素（Ⅰ类证据，A 级建议）。③目前不推荐：无选择使用抗凝剂、血液稀释疗法、多种神经保护剂等。国内指南给出了较细致的建议，例如：①对于不符合溶栓适应证且无禁忌证的缺血性脑卒中患者应在发病后尽早给予口服阿司匹林 150~300mg/d（Ⅰ级推荐，A 级证据）。②对大多数急性缺血性脑卒中患者，不推荐无选择地早期进行抗凝治疗（Ⅰ级推荐，A 级证据）。

所查到的指南和系统评价报告已对证据进行了质量评价，故可以直接进入到下一步。在没有相应指南及系统综述的情况下，应对纳入的各研究进行仔细评价，其中随机对照试验是一种前瞻性研究，在各种临床疗效研究中的论证强度最高，所以对大型随机对照试验也应给予一定的重视，明确各研究纳入的患者特征、各组干预措施的选择，从组间基线可比性、随访时间、随机化分组、盲法、失访率等方面评价各研究。

五、实践决策

经过研究结果的整理和分析，应用于具体病人的处理上。针对该病例，可以参考上述指南的推荐意见。故医生可以参照指南，根据当地的习惯和经验进行处理：①可以考虑静脉使用 rt-PA。②若医生及病人均不愿冒早期颅内出血增加的危险以获得远期残疾率降低的效益，也可不用溶栓治疗而立即使用阿司匹林、对症、支持和防止并发症及复发等措施，有条件则进入卒中单元。③病人的血压不是太高，故暂时不需使用降压药。医生将这些意见告知病人及家属，考虑到价格昂贵和担心出血的风险，病人方面选择放弃溶栓治疗，而选择了证据确凿的其他疗法。

六、小 结

急性脑梗死一种严重损害人类健康的常见病，脑血栓与脑栓塞所致的脑梗死占急性脑血管病的 70%，因此脑缺血性梗死的治疗越来越为人们所重视。本节具体应用循证实践的 5 个步骤，对急性脑梗死患者开展循证治疗。第一步，提出需解决的实际问题。第二步，检索相关证据，重点在于设计恰当检索策略，通过国内外各大权威数据库，全面检索出可信度高的证据。第三步，评价证据，根据适当的标准，对研究的质量特征进行严格评价，明确高质量研究。第四步，制定决策，根据当前最好的证据，结合患者的实际情况及意愿，给予患者最好的临床治疗。第五步，后效估计，即治疗决策应用后患者对治疗的反应。治疗指南及相关的系统综述无疑是疗效评价的最佳证据，但缺乏相应的证据时，临床医生就应对现有的随机对照试验等研究进行评价，结合患者的实际情况及治疗措施的利弊进行综合考虑，做出决策。

（时景璞）

笔记

第十一章　疾病预后的研究、评价与循证实践

本章节介绍了预后的基本概念、预后研究设计方案、基本步骤、主要统计分析方法、常见偏倚及其处理以及预后研究的基本评价原则和标准。预后是指疾病发生后，对将来发展为各种不同后果的预测或事前估计。预后研究包括研究疾病各种结局发生的概率以及影响预后的预后因素研究。预后研究最常用的研究设计方案是队列研究，包括回顾性队列和前瞻性队列研究。预后研究常用的观察指标有生存率、中位生存时间等，常用的统计分析方法有单因素 Log-rank 分析和多因素 COX 回归分析。对有关疾病预后研究的质量及其研究结论是否真实可靠，其评价的原则和标准可归纳为九条。

第一节　疾病预后的概念

一、疾病预后及其研究的意义

预后（prognosis）是指疾病发生后，对将来发展为各种不同后果（痊愈、复发、恶化、伤残、并发症和死亡等）的预测或事前估计，通常以概率表示，如治愈率、复发率、5 年生存率等。预后研究就是关于疾病各种结局发生的概率及其影响因素的研究。医生、患者及其家属都迫切需要了解该病的预后情况，医生知道该病预后情况，不仅对选择治疗方案有重要意义，而且可以回答患者及其家属所提出的各种问题。然而，要对预后做出客观估计与判断，尽可能使预后的结果接近患者的实际结局，有时有一定难度，只有进行疾病预后研究，在掌握了大量的预后信息后才能做到科学的预测。

疾病预后研究的意义在于：了解某种疾病的发展趋势和后果，从而帮助临床医师做出治疗决策；研究影响疾病预后的各种因素，有助于干预并改善疾病的预后。此外，可以从疾病预后研究中，来正确评定某项治疗措施的效果，从而促进治疗水平的提高。例如治疗癌症方案 A 的 5 年生存率高于方案 B，则说明方案 A 的疗效较好，应该选择方案 A。因此，疾病预后的研究具有重要的临床意义。

二、疾病的自然病史

疾病的自然病史（natural history）是指在不给任何治疗或干预措施的情况下，疾病从发生、发展到结局的整个过程。疾病的自然病史包括四个时期：

1. **生物学发病期**（biologic onset）　指病原体或致病因素作用于人体引起有关脏器的生物学反应，以及复杂的病理生理学改变，但此时很难用一般临床检查手段发现疾病已经发生。

2. **亚临床期**（subclinical stage）　是指病变的脏器损害加重，出现了临床前期的改变，患者虽无明显症状，自觉"健康"，但若采用某些实验室检查或特异性高及灵敏度高的诊断手段，可以做到早期发现、早期诊断和早期治疗。

3. **临床期**（clinical stage）　指患者病变脏器更加严重而出现解剖上的改变和功能障碍，临床症状、体征和实验室检查出现异常，因患者主动就医而被诊治。

4. **结局**（outcome）　指疾病经历了上述过程，发展到终末的结局，如痊愈、伤残或死亡等。

不同疾病，其自然病史差别很大，某些疾病自然病史较短，如急性感染性疾病，短期内出现症状体征和实验室异常，进展较快，较短时期内即可出现结局。而某些慢性非传染性疾病的自

笔记

然病史较长，甚至可达数十年之久，如心脑血管疾病、糖尿病、COPD 等，这些疾病的自然病史也比较复杂。研究疾病的自然病史对病因和预后研究、早期诊断和预防、判断治疗效果都有重要的意义。

三、临床病程

临床病程（clinical course）是指疾病的临床期，即首次出现症状和体征，一直到最后结局所经历的全过程，其中可经历各种不同医疗干预措施。临床医师可采取医疗干预措施来改变其病程。

病程的概念和疾病自然病史不同，病程可以因医疗干预（包括各种治疗措施）而发生改变，进而改变预后。在病程早期就采取积极医疗干预措施，往往预后较好，在病程晚期进行医疗干预措施的效果就不那么明显，疾病预后就比较差，因此，临床医师十分重视临床病程的估计。

四、预后因素

凡影响疾病预后的因素都可称预后因素（prognostic factors），若患者具有这些影响因素，其疾病发展过程中出现某种结局的概率就可能由此发生改变。预后因素的研究有助于临床医师及早进行医学干预，包括筛检、及时诊断、积极治疗和改变患者影响健康的不良行为等，从而改善患者的预后。预后因素和危险因素不同，危险因素（risk factor）是指作用于健康人，能增加患病危险性的因素，而预后因素是在已经患病的病人中研究与疾病结局有关的因素，因此，疾病的危险因素和预后因素是不同的概念。虽然有些疾病中某些因素既是危险因素，也是预后因素，但多数是不相同的，例如从图 11-1 可见急性心肌梗死的危险因素与预后因素，有些因素是相同的，且作用相似，如年龄和吸烟，随年龄增大，患病危险性增加，预后也差。但有些因素是相反的，如性别，男性发生急性心肌梗死危险性比女性增加，但发生心肌梗死后女性的预后比男性差。又如血压，高血压是危险因素，发生急性心肌梗死后低血压则预后不佳。

健康 ⟹ 急性心肌梗死发作 ⟹ 结局：死亡，再次梗死，恢复

危险因素：
- 高龄
- 男性
- 吸烟
- 高血压
- 高LDL/低HDL
- 不爱运动
- 炎症
- 凝血功能异常

预后不良因素：
- 高龄
- 女性
- 吸烟
- 低血压
- 前壁心梗
- 充血性心力衰竭
- 室性心律失常

图 11-1 急性心肌梗死的危险因素和预后因素之间的差别

影响疾病预后的因素是复杂多样的，概括起来有以下几个方面：

（一）早期诊断、及时治疗

任何疾病能否得到早期正确诊断，及时合理治疗，是影响预后的重要因素。尤其是恶性实体瘤，如能早期及时诊断，通过手术治疗，常能获得较好的预后。而发现较晚，已多处转移，失去手术根治机会，则预后很差。如胃癌，通过胃镜发现的早期胃癌，如微小胃癌术后 5 年生存率可达 100%，原位癌术后 10 年生存率也可达到 80%，如侵及黏膜下者，10 年生存率 65%，如侵及固有肌层者术后 5 年生存率 70%；而中晚期胃癌术后 5 年生存率仅为 16.9%。

(二) 疾病本身的特点

疾病本身的特点包括疾病的性质、病程、临床类型与病变程度等常是影响疾病预后的重要因素。某些自限性疾病如上呼吸道病毒感染，不需要治疗也可自愈，预后良好，同样是病毒感染如艾滋病和重症肝炎，预后就很差；败血症虽然病情很重，但可采用有效抗生素治疗而痊愈，但运动神经元疾病肌萎缩侧束硬化虽发展缓慢，但无有效治疗，预后很差，最终都因呼吸麻痹并发肺部感染死亡；霍奇金病的预后和病理类型有关，结节硬化型预后最好，5 年生存率 90% 以上，而淋巴细胞削减型预后最差，5 年生存率仅约 30%。

(三) 患者的病情

通常病情与预后密切相关，病情重者，预后较差。例如，黄疸腹水型重症传染性肝炎的预后远比无腹水的、轻或中型肝炎的预后为差。

(四) 患者的身体素质

患者的身体素质是一项综合指标，包括年龄、性别、营养状况、免疫功能等。同一种疾病，由于患者身体素质不同，预后差别可以很大。例如同一病理类型的非霍奇金淋巴瘤，如患者身体素质较差，年龄大，营养状况差，不能耐受强化疗，因而，病情易加重，生存期短，预后差。而身体素质好的患者，经过正规强化疗，不仅可长期生存，甚至可以治愈。

(五) 医疗条件

医疗条件的优劣，也会影响疾病预后。例如败血症可因抗生素选择不合理，疗效差；如果结合细菌培养、药物敏感试验合理选用抗生素，疗效可以提高，预后较好。又如急性心肌梗死在医疗条件差的医院，许多有效治疗措施都不能实施，病死率较高；而条件好的医院不仅医疗设施好，患者早期的正确诊断率高且有抢救经验丰富的专科医师及许多有效治疗措施如溶栓治疗、经皮冠状动脉腔内成形术、冠状动脉支架术、冠状动脉搭桥手术等都可以选择，从而可以降低病死率，改善预后。

(六) 社会、经济和家庭因素

如医疗制度、社会保险制度、家庭成员之间关系、家庭经济状况、文化教养、患者宗教信仰及心理因素等也会影响患者疾病的预后。

第二节　疾病预后研究中常用的结局指标

一、各种率的指标

(一) 病死率(case-fatality rate)

在患某病患者总人数中，死于该病的患者所占的比例。常用于病程短且容易死亡的疾病，如各种传染病、急性中毒、心脑血管疾病的急性期和迅速致死的癌症。

病死率 =(因该病死亡人数 / 患该病接受治疗的总患者人数)× 100%

(二) 疾病死亡率(disease-specific mortality)

一定的时期内(通常指一年)，某一人群中因为某病死亡的人数所占的比例，一般以 1/10 万或 1/ 万为单位。

(三) 治愈率(cure rate)

系指患病治愈的患者人数占该病接受治疗患者总数的比例。

治愈率 =(患某病治愈的患者人数 / 患该病接受治疗的总患者人数)× 100%

(四) 缓解率(remission rate)

进行某种治疗后，进入疾病临床消失期的病例数占总治疗例数的百分比。有完全缓解率、部分缓解率和自发缓解率之分。

笔记

缓解率 =（治疗后进入疾病临床消失期的病例数 / 接受该种治疗的总病例数）× 100%

（五）复发率（recurrence rate）

疾病经过治疗已有一定程度缓解或痊愈后又重复发作的患者数占观察患者总数的百分比。

复发率 =（复发的患者数 / 接受观察的患者总数）× 100%

（六）总体生存率（overall survival rate，OS）

从疾病临床过程的某一点开始，能存活一段时间的病例数占总观察例数的百分比。生存率常用于病程长、致死性疾病，如各种癌症。病程较短的可用 1 年生存率，较长者用 5 年生存率表示预后。例如急性髓系白血病（AML）的 5 年生存率为 20%，表明从诊断 AML 开始随访有大约 20% 的患者可以生存 5 年以上。

预后研究仅仅报道生存率是不够的，例如图 11-2 中 4 种情况 5 年生存率均为 10%，但 5 年间生存率的变化趋势截然不同。

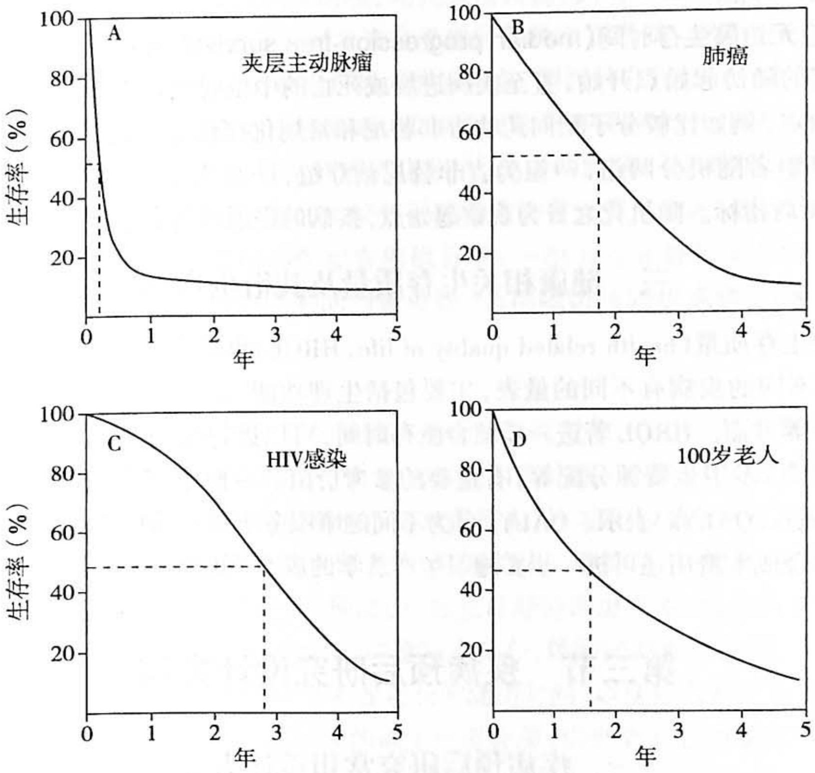


图 11-2　四种不同人群的生存曲线

注：5 年生存率均为 10%，但生存曲线明显不同，说明仅报道 5 年生存率有很大的局限性，虚线为中位生存时间

（七）无病生存率（disease-free survival rate）

常用于癌症的结局判断，指疾病经过治疗达到临床缓解后，未出现临床疾病复发或死亡的患者占所有临床缓解患者的比例。例如 100 例 AML 患者经过某一化疗方案治疗，有 70 例达到完全缓解，3 年内有 20 例复发，各种原因死亡 5 例，则应用该化疗方案治疗 AML 的 3 年无病生存率为 45%（45/100）。

（八）无进展生存率（progression-free survival rate，PFS）

常用于癌症的结局判断，指疾病经过治疗达到病情稳定后，没有临床疾病进展或死亡的患者占所有临床稳定患者的比例。例如 100 例肺癌患者经过手术治疗，达到病情稳定，3 年内 75 例病情无进展，20 例病情有进展，各种原因死亡 5 例，则 3 年无进展生存率为 75%（75/100）。

笔记

二、中 位 时 间

（一）中位生存时间（median survival time）

又称为半数生存期，即当累积生存率为 0.5 时所对应的生存时间，表示只有 50% 的患者可以活过这个时间。例如 AML 的中位生存期为 20 个月，说明诊断为 AML 后只有 50% 的患者可以活过 20 个月。

利用生存曲线可以计算中位生存时间，图 11-2 中夹层主动脉瘤的中位生存时间不足半年，肺癌不足 2 年，HIV 感染约为 3 年。

（二）中位无病生存时间（median disease-free survival time）

基本概念同上，如果 AML 的中位无病生存时间为 18 个月，表明有 50% 的患者可以在无临床疾病的状态下存活 18 个月。无病生存时间一般从疾病缓解之日、手术切除之日开始算起，到疾病复发或死亡为止。

（三）中位无进展生存时间（median progression-free survival time）

指从规定的随访起始点开始，直至疾病进展或死亡的中位时间。在疗效评价中，常用随机化之日为起始点，例如比较分子靶向药物吉非替尼和常规化疗作为一线治疗肺癌的长期疗效，新诊断的肺癌患者随机分两组，一组为吉非替尼治疗组，一组为常规化疗组，长期随访，观察 PFS 和 OS 等预后指标。随机化之日为观察起始点，疾病的进展或各种原因的死亡为观察终点。

三、健康相关生存质量及其衍生指标

健康相关生存质量（health related quality of life，HRQL）也是常用的预后研究指标，常用量表进行评定，不同的疾病有不同的量表，主要包括生理功能、心理功能、社会功能和对健康状况的总体感受等方面。HRQL 若进一步结合生存时间，可以进行效用评价。效用指标对于临床医疗、卫生决策以及卫生资源分配等，有重要的参考价值。一般采用质量调整寿命年（quality-adjusted life years，QALYs）表示。QALYs 可为不同健康服务干预措施的结果评价与抉择提供重要信息。若结合成本费用还可进一步实施卫生经济学的成本 - 效用分析。

第三节 疾病预后研究设计方案

一、疾病预后研究常用设计方案

疾病预后研究包括预后因素的研究及预后的评定。基于研究目的和可行性，可以选择描述性研究、病例对照研究、回顾性队列研究、前瞻性队列研究等不同设计方案。理论上最佳的设计方案是随机对照临床试验，但其可行性差，所以预后研究最常用的研究方案是队列研究，包括回顾性队列研究和前瞻性队列研究，以后者为佳。研究设计方案不同，其研究结果可以相差很大。例如，研究方法不当造成泌尿系统结石的复发率可由 20% 到 100% 不等，不同研究设计的溃疡性结肠炎癌变率结果可从 3% 至 10%，相差数倍。

疾病预后的评定指标，如病死率、治愈率、缓解率、复发率、致残率、生存率等，可以通过长期随访研究对象的纵向描述性研究获得。若要比较两组病例的预后，如两组生存率比较等，可以借助纵向的队列研究完成。

二、疾病预后研究设计中的若干注意事项

（一）队列研究的起始点

笔记

预后研究若采用队列研究设计，首先应确定队列研究的起始点，又称为零点时间（zero time），

该起始点在研究设计时必须要明确规定，是从病程的哪一点开始观察，两个队列中的每一个研究对象都要使用同一起始点，进行追踪和观察以及预后结局的比较。对于预后研究，要尽可能选择疾病的早期，若队列的集合时间接近疾病初发时点，则称起始队列（inception cohort），为队列研究的首选。但在队列研究中，病人往往处于病程的不同阶段，开始观察时间不同，很难确定康复、复发、死亡等具体时间，倘若有些病人采用普查检出时间作为零点时间，有的是按治疗开始日算起，零点时间不统一，难以评价真正的预后，甚至得出错误的结论。例如，同一组 AML 患者，如果以诊断之日作为研究起始点，其中位生存时间为 20 个月，如果以化疗第一天作为起始点，则中位生存时间为 28 个月，因为有一部分患者诊断后还没有来得及化疗就已死亡，未被纳入研究，造成结果高估。而以诊断之日作为起始点，则更为合理。

（二）研究对象的来源和分组

研究对象要具有代表性，能代表目标疾病的人群。若病人来自不同级别医院，其疾病预后可能不同。如果研究对象全部来自三级医院，则代表性较差，因为三级医院的患者往往病情较重且接近病程后期，预后差。再如研究对象来自某地区各种级别医院，通常包括了各种型别及其病情严重程度各异的病例，更能反映出目标人群的特点，因而具有较好的代表性。

研究对象的分组也必须遵循可比性原则，即非研究因素在组间的分布是均衡可比的。在评价治疗方案对预后的影响时，尽量做到随机化分组，确保两组除了拟研究的治疗方案外，其他各种影响预后的因素能够相同。例如比较分子靶向药物吉非替尼和常规化疗作为一线治疗肺癌的长期疗效，应该对新诊断的肺癌患者随机分组，一组为吉非替尼治疗组，一组为常规化疗组，保证两组病情特征等预后因素的均衡可比，长期随访并同步观察 DFS 和 OS 等预后指标。

（三）随访和失访

预后研究中随访工作十分重要，要尽量随访到所有研究对象，失访率越低越好。如失访率小于 5% 一般可以接受，对结果的影响小，大于 10% 应引起注意，若超过 20%，则研究结果的价值不大，这是因为患者失访会造成疾病预后信息大量丢失，影响预后结果的可靠性。减少病例失访应注意以下数点：加强对患者及其家属宣传、强调随访的重要性，以提高随访的依从性；建立健全随访管理制度，做到专人负责、定期随访；建立良好的医患关系，做到有求必应，不失信于患者；以及改进随访方式、体现人文关怀，尽量采用关心、体贴的调查语言等。

随访期限视疾病病程而定，原则上要有足够长的随访时间，以便能观察到疾病的所有结局，包括一些罕发的不良反应。随访间隔时间的确定也要合理，以便能动态观察到各种变化情况。随访间期不同，可导致不同结论。例如同样一组肺癌患者队列，如果观察间期分别定为 2 个月和 6 个月，间隔 2 个月的随访可能更容易被观察到肿瘤的进展情况，DFS 时间可能短于 6 个月的患者。一般对于病程短的疾病，随访间隔时间可短些，对于病程长的疾病，随访间隔时间可适当延长。随访过程中确定的各种结局，一定要有明确的定义和判断标准，在设计时就应规定好，执行中不再变动，标准要有客观性；最好用盲法观察，以防止测量偏倚。

对失访者的处理可采用两种简单的方法，一是按死亡计，或者直接从观察患者人数中删除，不予统计，这两种方法均损失预后的信息，最好是按照本章第四节介绍的生存分析方法，处理失访病例，以减少信息损失。对有失访病例的研究，可采用敏感性分析估计失访对研究结果的影响，具体方法为先假定失访者均出现预定结局，得到结局事件“最高”发生率，然后假定失访者均不出现结局事件，得到“最低”发生率，比较“最高”和“最低”率，如两者相差不大，则结果可取，如两者相差很大，则研究结果不可靠。例如，有 100 例研究对象，研究结束时 4 例死亡，16 例失访，粗病死率为 4.8%（4/84），按照最坏估计，16 例失访者均死亡，则最高病死率为 20%（20/100），按照最好估计，16 例失访者均存活，则最低病死率为 4%（4/100），最高病死率 20% 与粗病死率 4.8% 相差很大，认为失访对研究结论有比较大的影响，该研究结果不可信。

笔记

三、疾病预后研究的基本步骤

以上海市中美白血病协作组完成的623例AML的预后研究为例，说明预后研究的基本步骤。

1. 确立研究的疾病、研究目的　研究原发性、初治AML的预后。

2. 确定观察指标、研究因素　结局观察指标为完全缓解率，3年复发率，3年生存率，中位生存时间，分析影响生存的预后因素。

3. 确定研究方案　采用前瞻性队列研究的设计方案进行预后研究。确定AML的诊断标准，定义OS为从诊断之日到死亡或最后的随访之日，确定随访终止时间为2008年5月1日。

4. 收集队列信息、登记资料　前瞻性收集2003—2007年的上海市白血病协作组24家医院的原发性、未治疗的AML患者623例，所有骨髓标本统一送到中美实验室进行统一诊断，以防误诊。登记病例信息和所有相关的实验室资料，包括身份证号码、联系电话、地址、主管医师的联系信息等。

5. 随访　每6个月随访一次，同时随访患者本人(或家属)和主管医师，随访内容包括症状、体征、治疗方案、并发症、血象和骨髓检查结果。多种途径随访对减少失访有很大帮助。

6. 统计分析(详见本章第四节内容)

(1)病例一般资料分析：患者年龄、性别、WHO亚型、染色体等资料的分析。

(2)完全缓解率、复发率的计算。

(3)生存率、中位生存时间、生存曲线：应用Kaplan-Meier方法计算生存率，绘制生存曲线。

(4)单因素分析：应用Log-rank方法统计与生存相关的预后因素，得出与预后相关的因素如WHO亚型、染色体分组等。

(5)多因素分析：把单因素分析中$P<0.1$的预后因素纳入COX风险比例模型进行多因素预后分析。

第四节　疾病预后研究统计学方法

一、生存率和中位生存时间分析

(一)生存率的计算

在生存分析中，生存率的计算有两种方法：直接法和间接法。

1. 直接法　如果病例数多，没有失访，则结果可靠，计算简单。例如5年生存率=活满5年的例数/总随访例数。但一般生存数据均存在删失值(censored data)，也称为终检值、截尾数据，需要用间接法计算生存率。

2. 间接法　删失值包括三种情况：①失访；②死于其他疾病；③观察到规定的随访截止时间仍存活。常用的间接法包括Kaplan-Meier法和寿命表法。其中，寿命表法用于例数较多的研究，Kaplan-Meier法可用于小样本研究，也可用于大样本的研究。目前有多种统计软件可以进行生存分析，不需手工计算。

生存曲线是以随访时间为横坐标，累积生存率为纵坐标作图即为生存曲线，常用的Kaplan-Meier曲线，表示时间与生存关系的函数曲线。随访时间单位越小，则精密度越高，即生存期用月计比年计为佳。生存曲线分析能获得有关疾病过程任何时刻的生存率，信息量远远超过点估计值。例如几种疾病的5年生存率相同，但生存曲线却大不相同(图11-2)。小样本获得的曲线为阶梯形，而大样本就形成光滑的曲线。应用间接法进行生存分析时必须注意，所绘生存率曲线纵坐标所示是一个假想队列的生存概率，而不是患者实际的生存率，因左侧的观察病例数总比右侧多，因此曲线左侧的估计值较右侧可靠。

笔记

生存资料需要具备以下几个要素才能进行生存分析：结局、随访开始和结束时间、研究因素。按照表 11-1 数据格式整理后，可便于统计分析。随访开始和死亡时间最好要精确到天，例如最后随访截止日期为 2009 年 12 月 31 日（表 11-1）。多数统计软件进行生存分析时默认事件发生为 1，未发生为 0。

表 11-1　生存分析数据整理格式

病例编号	诊断	随访起始日期	死亡或截止日期	结局(死亡=1)	变量 1	变量 2
1	1	2000/04/05	2005/05/06	1	1	2
2	2	2004/03/02	2009/12/31	0	4	1
3	1	2005/07/05	2009/12/31	0	3	1
4	3	2003/06/01	2008/08/09	1	1	2
5	4	2002/08/10	2009/12/31	0	4	2
……						

（二）中位生存时间的计算

预后研究中还需要计算生存期的长短，即存活期。常用总体中位生存期、无病存活期（disease free survival，DFS）、无事件存活期（event free survival，EFS）、无进展存活期（progression free survival，PFS）等来表示。由于生存时间是一种呈正偏态的连续变量，因此要计算中位生存期，不能计算平均生存期；鉴于删失值的存在，只能用生存分析计算中位生存时间。生存分析方法可以处理截尾数据。

例如，有 12 例患者，8 例死亡，2 例失访，2 例截止到随访日尚存活，生存时间分别为 10、20、20、30、80、50、100、150、80^+、10^+、120^+、200^+ 个月，按照生存分析计算中位生存时间为 80 个月，而用传统方法计算该组数据的中位数为 65 个月。

二、疾病预后因素的研究方法

（一）单因素预后因素分析

从前述已知，影响疾病预后的因素很多，包括患者一般情况如性别、年龄、体质和营养情况、社会经济和心理状况等，疾病本身的情况如病理组织学类型、病灶大小、病原体种类、临床分期等，治疗方法及患者、医护人员的依从性等。对疾病预后因素的识别和研究，是疾病预后研究中另一项重要内容。预后因素的研究方法和疾病危险因素的研究方法相似。一般可以先从回顾性的临床资料中进行筛检，然后通过病例对照研究、前瞻性队列研究进一步加以论证，从而确定是否为预后因素。分析方法可先从单因素分析开始，然后进行多因素分析校正各种混杂因素，一般把单因素分析中 $P<0.1$ 的预后因素纳入多因素预后分析。

单因素分析常用的统计方法是 Log-rank 分析，并以 Kaplan-Meier 生存曲线的形式来直观地展示，上述 12 例患者男性和女性的预后见表 11-2 和图 11-3，说明女性预后良好。单因素分析中 $P<0.1$ 的因素均应纳入多因素分析。

表 11-2　采用 Log-rank 分析进行单因素预后分析

		例数	死亡例数	中位生存时间(个月)	χ^2 值	P 值
性别	男	7	5	30	6.55	0.01
	女	5	2	150		

在单因素研究中确定某因素是否系预后因素时，必须保证观察组（存在某预后因素）和对照组（不存在该预后因素）两组的临床特点和其他非研究因素都要相同，但在实际工作中常不易做到。所以单因素分析结果提示某个有意义的预后因素很可能不是独立的预后因素。为了尽量

笔记

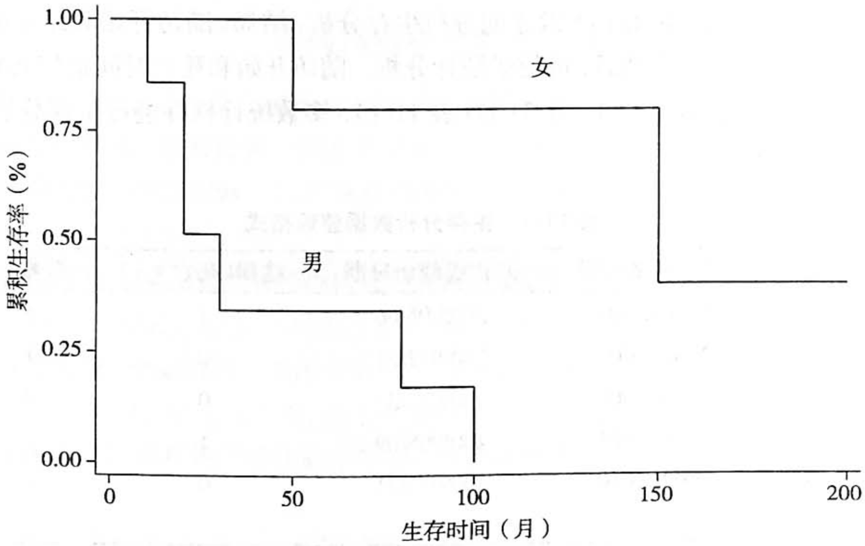


图 11-3 不同性别患者的 Kaplan-Meier 生存曲线

减少混杂性偏倚，可以采用下列方法，如限制、配比、分层及标准化等方法加以平衡。如有多个混杂因素，应该进行多因素分析。

（二）多因素预后因素分析

由于疾病的结局和多种预后因素有关，各种预后因素又是互相影响的，为了全面正确地衡量预后因素的作用，可以采用多元回归、逐步回归、logistic 回归及 Cox 模型等多因素分析方法，进一步筛选出与疾病结局有关的主要预后因素，以建立该疾病预后函数或预后指数。1972 年英国统计学家 D.R. Cox 提出一种能处理多因素生存资料数据的回归模型，称风险比例模型（proportional hazard model），简称 Cox 回归或 Cox 模型。在队列研究中应用最广，可以允许终检值即截尾数据的存在，终检的原因可能是死于其他原因、失访或到资料总结时随访对象还活着但尚未发生所规定的事件。同时 Cox 模型还能有效处理随访时间长短不一及资料失访等预后研究中经常碰到的问题，难以用常规统计方法解决。在 Cox 模型中，在时点 t，除了有一个基本（本底）风险量 $h_0(t)$ 外，第 i 个影响因素使该基本量 $h_0(t)$ 增至 $e\beta_iX_i$ 倍而成为 $h_0(t)e\beta_iX_i$。因此，如果 k 个因素同时影响生存过程，在时点 t 的风险量是 $h_0(t)\times e\beta_1X_1\times e\beta_2X_2\times\cdots\times e\beta_kX_k$。所以 Cox 模型的基本结构为：

$$h(t)=h_0(t)\times e\beta_1X_1\times e\beta_2X_2\times\cdots\times e\beta_kX_k \qquad 式(11\text{-}1)$$

以死亡事件为例，则公式中 $h(t)$ 表示在时点 t 的死亡率，$h_0(t)$ 表示在时点 t 的本底死亡率，可理解为在时点 t 时不存在 $X_1, X_2, \cdots, X_k$ 的影响情况下的死亡率（即 $\beta_1=\beta_2=\cdots=\beta_k=0$ 时的死亡率），$h_0(t)\times e\beta_1X_1$ 表示第 1 个因素 X_1 所产生的作用，使死亡率 $h_0(t)$ 增至 $h_0(t)\times e\beta_1X_1$，β_1 可理解为 X_1 的回归系数。

在 Cox 模型作分析时，可估计风险比（hazard ratio，HR），$HR=e^{\beta}$，其意义与相对危险度（RR）相同。表 11-3 是有关 435 例骨髓增生异常综合征（MDS）的预后因素分析，年龄≥60 岁、中性粒

表 11-3 骨髓增生异常综合征 Cox 模型多因素预后分析

因素	回归系数 β	HR(95% CI)	z 值	P 值
年龄（≥60 岁）	0.63	1.87（1.38~2.54）	4.00	0.000
中性粒细胞绝对值（$<1.0\times10^9$/L）	0.37	1.45（1.05~2.01）	2.24	0.025
血红蛋白（<90g/L）	0.87	2.38（1.50~3.79）	3.66	0.000
血小板 $<30\times10^9$/L	0.36	1.44（1.06~1.95）	2.34	0.019
IPSS 积分分组	0.74	2.10（1.71~2.56）	7.16	0.000
WPSS 积分分组	0.19	1.22（0.90~1.50）	1.92	0.055

笔记

细胞绝对值 $<1.0\times10^{9}/L$、血红蛋白 <90g/L、血小板 $<30\times10^{9}/L$、国际预后积分（IPSS）高危组均是影响 MDS 生存的独立因素，例如血红蛋白 <90g/L 的患者与血红蛋白≥90g/L 的患者相比，在校正了其他影响因素后，死亡风险增加了2.38倍。

第五节　预后研究中常见的偏倚及其处理方法

一、预后研究中常见的偏倚

（一）集合偏倚（assembly bias）

集合偏倚或称分组偏倚、就诊偏倚、易感性偏倚（susceptibility bias），属于选择性偏倚。由于各医院的性质和任务不同，收治的患者在病情、病程和临床类型可能不同；就诊患者的经济条件在不同地区也可能有所不同，影响了疾病的早期诊治。在集合队列进行随访，随访结束时发现预后不同可能并非研究因素所致。

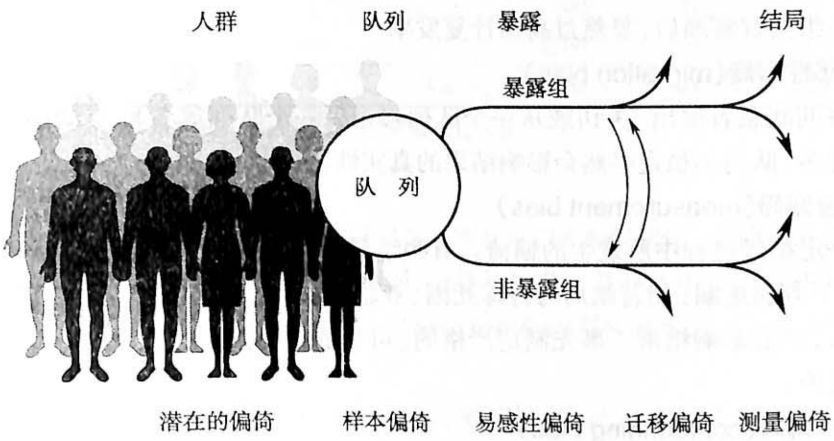


图 11-4　队列研究中常见的偏倚

（二）存活队列偏倚（survival cohorts bias）

从各医院收集病例组成队列进行预后研究，并非都是起始队列（inception cohort），而是该病病程中某一时点进入的存活病例，故称存活队列偏倚（图 11-5），那些未入院的失访病例，因信息丢失，造成预后失真。比如心肌梗死，可能出现入院前死亡，造成实际入院病例的病情较轻，从而高估预后。若采用起始队列进行研究，集合队列 150 例，随访结果时，预后好的 75 例，预

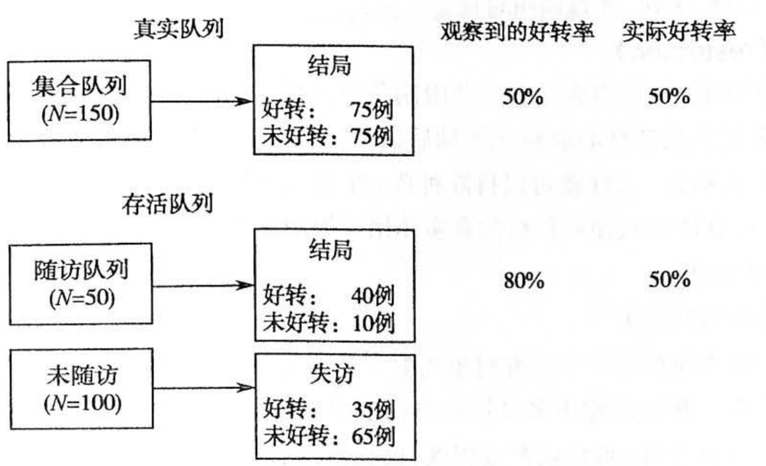


图 11-5　存活队列偏倚

后不好的 75 例，各占 50%；若以医院存活病例组成队列，假如有 50 例，其中预后好的 40 例，预后差的 10 例（20%），前者占 80%（40/50），而未被随访到的 100 例中预后好的 35 例，预后不佳 65 例，占 65%（65/100），估计的预后不良率分别为 20% 和 65%，而真值却为 50%，出现严重偏差。因此，存活队列偏倚实际上也是集合偏倚的一种特殊类型。

（三）失访偏倚（lost to follow-up）

由于观察时间长，观察对象因迁移、外出、不愿继续合作、因药物不良反应而停止治疗或死于非终点疾病等原因脱离了观察，即失访，造成的偏差，称失访偏倚，属于选择性偏倚，在预后研究中尤为常见。如一项预后研究中，有 100 例病人进入随访观察，但最后有 20% 病人失访，其中随访完整的 80 例病例的疗效为 80%。倘若失访 20 例的疗效也为 80%，那么该研究的实际疗效为 80%，相反若 20 例中有效的仅占 20%，则实际疗效为 68%。可见失访偏倚可使研究结果严重失真。

（四）零时不当偏倚

若研究对象开始观察的起始时刻不一致，分别处在该疾病病程的不同阶段而产生的偏差，称零时不当偏倚。如随访观察肾结石的复发率，住院病人由于观察的零时刻不同，若将初发与复发混在同一组去观察预后，显然过高估计复发率。

（五）迁移性偏倚（migration bias）

随访观察期间患者退出、失访或从一个队列移至另一个队列等各种变动引起的一种偏倚。变动的人数过多、队列不稳定必然会影响结果的真实性。

（六）测量偏倚（measurement bias）

观察与评定结局过程中所发生的偏倚。有些结局指标，如死亡、脑血管意外、某些肿瘤，诊断十分明确，不容易遗漏；倘若结局为特殊死因、亚临床疾病、不良反应、残疾等就不那么清楚，判断或有出入，从而影响结果。事先制定严格的、可行的结局判断标准，采用盲法评定结局，可以减少测量偏倚。

（七）混杂偏倚（confounding bias）

在研究某因素是否为预后因素时，理论上必须保证观察组（存在某预后因素）和对照组（不存在该预后因素）两组的临床特点和其他非研究因素都要相同，但在实际工作中常难以做到。

二、偏倚的处理方法

（一）随机化（randomization）

从理论上讲，两个队列进行比较，应当除研究的预后因素外，其他因素最好组间均衡，即基线状况要相同。随机化分组是消除选择性偏倚最好的方法，可使每个研究对象都有同等的机会进入观察队列和对照队列，实现两组可比。

（二）限制（restriction）

在选择研究对象时，可事先设置一些限制条件，以排除其他一些已知影响因素对结果的干扰。如研究年龄是否系急性心肌梗死的预后因素，可将研究对象限制在黄种人、男性、无并发症的前壁心肌梗死病人，这样就可以排除种族、性别、心肌梗死部位和并发症等因素的干扰和影响，分析年龄对急性心肌梗死预后的真实作用。但用此法控制偏倚，其预后结果常有很大的局限性，结论外推性差。

（三）配比（matching）

配比就是为观察组的每一个研究对象匹配一个或几个类似的对照，旨在消除这些配比因素对结果的潜在影响。预后研究中多以年龄 / 性别和种族作为配比因素，因为这些因素也是最常见的混杂因素。除此之外，像疾病严重程度、既往治疗史等也用来配对。但千万不能把研究因素作为配对条件。

笔记

（四）分层（stratification）

分层常是最常用的检出和控制偏倚方法，特别是有潜在的混杂偏倚时，应用分层方法可以在研究设计和资料分析阶段有效控制偏倚。例如 A、B 两个医院冠心病搭桥术后病死率比较（表 11-4），A 医院粗病死率为 4%，高于 B 医院的 2.6%，认为 A 医院手术风险高。但如果以术前危险度进行分层，重新计算各层内的标化死亡率，A 和 B 医院又是相同的，为什么会出现这样的结果呢？主要是因为两所医院病人的病情构成不同，A 医院高危病例较多，而 B 医院低危病例较多，分层分析消除了病情构成这个混杂因素的影响。

表 11-4 分层分析案例

危险度	A医院				B医院			
	病人数	死亡数	病死率（%）	标化病死率（%）	病人数	死亡数	病死率（%）	标化病死率（%）
高	500	30	6		400	24	6	
中	400	16	4		800	32	4	
低	300	2	0.67		1200	8	0.67	
合计	1200	48	4	3.6	2400	64	2.6	3.6

注：两所医院冠心病搭桥术后病死率，以术前病情严重程度分层

（五）标准化（standardization）

当比较两个率时，若两组对象内部构成存在差别且足以影响结论时，可用标准化率校正。最简单的方法就是假设两组的内部构成或权重相同，计算加权率。表 11-4 中 A 医院和 B 医院的标化病死率 =（1/3×6%）+（1/3×4%）+（1/3×0.67%）=3.6%，每组的内部构成均为 1∶1∶1，即权重均为 1/3。

（六）多因素分析方法

在预后因素研究中，常有多个预后因素同时影响结局，此时应借助多因素分析方法，从中筛选出与疾病结局有关的主要预后因素，探讨这些因素对预后的作用大小。在预后因素研究中以比例风险模型 -Cox 回归分析方法最为常用。

第六节 疾病预后研究的评价原则

对疾病预后研究的评价，同样包括三个方面：真实性、重要性和适用性。具体评价原则和标准可归纳为以下九条（表 11-5）：

表 11-5 预后研究文献的评价原则

预后研究文献的评价原则
一、真实性
1. 队列的起始点是否相同？
2. 队列是否有代表性？
3. 随访是否足够长，是否完整？
4. 判断结局时是否有客观的结局标准，是否采用盲法？
5. 是否对影响预后研究的重要因素进行了统计学的校正？
二、重要性
1. 预后研究的结果报告是否完整？
2. 研究结果的精确性如何？即可信区间是否较窄？
三、适用性
1. 我们自己的病人是否与文献报道的病人非常不同？
2. 研究结果是否有助于治疗方案的制定和是否有助于对患者及其亲属作出解释？

笔记

一、观察预后的研究对象是否都处于同一起始队列

预后研究要求各队列研究对象的预后观察起始点一定要统一，可以是症状首发时间、疾病确诊时间或治疗开始时间，零点时间务必明确一致。如研究脑卒中的预后因素，纳入的研究对象应是首次发作的脑卒中患者，排除复发者。对入选的研究对象处于病程的哪一个阶段必须清楚叙述。零点时间最好是处于病程的早期，即起始队列（inception cohort）。如研究 AML 的生存时间，就是统一以诊断之日为零点时间，不能有的是诊断时间，而有的又是以开始化疗为起点时间。

二、研究对象是否能代表被研究疾病的目标人群

研究对象应有明确的诊断标准、纳入 / 排除标准。对研究对象的来源机构、来源地应作详细叙述，以便了解研究对象的代表性，判断是否存在选择性偏倚。对研究对象的一般特征包括年龄、性别、疾病严重度和有否并发症存在等也应详细介绍，因为这些因素都和预后有关。

三、随访时间是否足够？随访是否完整

随访时间必须足够长，以便发现关注的研究结果。如果随访时间过短，发生了结局事件（如肿瘤发生、康复、复发等）的患者中，只有一小部分被发现，难以反映该疾病预后的真实情况。

同时随访必须充分、完整。在理想情况下，所有纳入研究对象的观察应由始至终，即从疾病早期一直随访到完全康复、复发或死亡，但事实上难以做到，失访不能完全避免。如何来判断失访对结论的影响，一般遵从“5 和 20”原则，失访率 <5%，其研究结果可靠，如失访率 >20%，结果真实性将大打折扣，若失访率在 5%~20% 之间，则结果较可靠。亦可通过前述的敏感性分析，比较“最高”和“最低”发生率，如两者相差不大，则结果可信，如两者相差很大，则研究结果不可信。

四、判断结局有无客观标准，是否采用了盲法

观察疾病预后的终点，即结局，应有客观的标准。在研究开始前，研究者必须对结局有明确的定义，同时要有客观的测量标准。有些预后容易确定，如死亡，但大多数结局，如痊愈、残疾、复发、生存质量改变等，都需要有客观的标准，以避免判断预后结局时出现分歧，影响结论。若预后结局属“硬”指标，如“死亡”、“残疾”等可不用盲法判断，倘若结局属“软”指标，如一过性脑缺血、不稳定性心绞痛，则应采用盲法判断，以避免发生疑诊偏倚（diagnostic-suspicious bias）以及预期偏倚（expectation bias）等。

五、是否对影响预后研究的重要因素进行了统计学的校正

预后研究中可能存在各种混杂因素，应对这些因素进行统计学校正。Framingham 的研究者报道风湿性心脏病心房颤动病人的脑卒中发生率为 41/1000 人年，与非风湿性心脏病心房颤动病人的脑卒中发生率十分接近，但进一步研究发现风湿性心脏病患者的年龄普遍年轻。对患者的年龄、性别和高血压状态进行统计学校正后，风湿性心脏病心房颤动病人脑卒中的发生率是非风湿性心脏病心房颤动病人的 6 倍。

校正方法包括最简单的分层分析以及较复杂的诸如 logistic 回归及 Cox 模型等多因素分析法。

六、预后研究的结果报告是否完整

比如报告生存率有三种方法：①某一时间点的生存率，如 1 年生存率、5 年生存率等；②中位生存时间（median survival time），即观察到 50% 的研究对象死亡的随访时间；③生存曲线

笔记

(survival curves)可以了解预后可随时间而变化。完整的结果报告应当同时报告某一时点的生存率、中位生存时间以及生存曲线，否则可能出现挂一漏万的情况。例如，某研究中两组的 1 年生存率都是 20%，但二者的生存曲线形态可以完全不同，一条显示中位生存时间为 3 个月，提示疾病早期预后就很差，另一条显示中位生存时间为 9 个月，提示疾病早期预后好。

七、研究结果的精确性如何

除了报告生存率、生存时间、生存曲线，还应提供预后估计的精确度，即预后结局概率的 95% 可信区间。对预后因素的研究可用相对危险度和绝对危险度等来表示，同时也要报告 95% 可信区间。95% *CI* 较窄，说明样本量足够大，对总体预后的估计更精确。

八、我们自己的病人是否与文献报道的病人非常不同

临床上是否可以采纳文献报道的结果，还要关注我们的病人与文献报道的研究对象是否在年龄、性别、疾病特征等方面相似或存在极大的不同？如果差别不大，就可以考虑应用这些研究结果。

九、研究结果是否有助于治疗方案的制订和是否有助于对患者及其亲属作出解释

研究结果是否直接有助于治疗方案的取舍？例如在非风湿性心房颤动病人中应用华法林抗凝治疗，能降低缺血性脑卒中的发生率，但在另一项"孤立性心房颤动"病人(60 岁以下，无相关的心肺疾病)的研究中，15 年内脑卒中发生率仅为 1.3%，长期应用华法林可能弊大于利，该研究对确定华法林抗凝治疗的取舍有重要临床价值。

研究结果是否有助于对患者及其亲属作出解释？例如一项高质量的研究结果显示疾病的预后良好，则有助于向焦虑的病人及其家属作出解释而使其放心。另一方面，若一项质量高的研究结果显示疾病预后不良，也可与病人和其家属进行沟通，同样具有实用价值。

第七节　循证实践及案例分析

一、疾病案例

男性患者，40 岁，确诊为急性髓系白血病(AML)M4 亚型 1 个月，染色体正常核型，已经给予标准 DA 方案(柔红霉素 + 阿糖胞苷)诱导缓解(CR)，缓解后准备给予缓解后的巩固强化治疗。缓解后的巩固强化治疗非常重要，正确治疗可以减少复发，延长生存期。缓解后治疗可以采用多种标准剂量化疗药物组合的常规方案，也可以用大剂量阿糖胞苷(Ara-C)单药治疗，到底何种治疗方案更好，复发率低，生存期长呢？即哪种治疗方法可以提高 AML 患者的预后？

二、提出问题

采用 PICO 方式提出临床问题：对于 AML 患者而言，缓解后应用标准化疗方案和应用大剂量 Ara-C 化疗方案进行巩固强化治疗，复发率、生存率、生存时间何种治疗更佳？不良反应是否可以耐受？

三、证据检索与评价

首先以(acute myeloid leukemia)AND(high dose cytarabine)AND(overall survival)为检索式，限定在"Title/abstract"和"Meta analysis"，未检索到相关的 meta 分析。重新检索，把检索条件限

笔记

定在“Title/abstract”和“Randomized Controlled Trial”和近5年的文献，检索到20篇相关文献。阅读题目和摘要，选择最新的、样本量较大、多中心研究、随访时间较长、发表在影响因子较高的文献杂志，阅读全文，按照预后文献评价原则进行文献评价。

（一）文献基本内容

1. 文献题目 A randomized comparison of 4 courses of standard-dose multiagent chemotherapy versus 3 courses of high-dose cytarabine alone in postremissiontherapy for acute myeloid leukemia in adults: the JALSGAML201 Study.

2. 文献来源 Blood, 2011, 117: 2366-2372.

3. 研究目的 在年龄<65岁的AML患者获得第一次完全缓解（CR）后，比较2种治疗方案对患者预后的影响。一种方案为4个疗程的标准剂量、多种化疗药物组合的传统方案（CT），另一种方案为3个疗程的大剂量Ara-C（HiDAC）。

4. 研究设计 为随机对照临床试验，CONSORT流程图见图11-6。2001年12月~2005年12月期间共有1057例新诊断的AML符合纳入标准，年龄15~64岁，来自129个研究中心。第一次随机分组，分别用IDA + Ara-C和DNR + Ara-C诱导缓解方案进行诱导治疗，共有823例患者取得CR，CR后再次随机分组，分别用4个疗程的大剂量Ara-C和3个疗程的标准传统方案，42例患者因为各种原因未参加第二次随访分组。中位随访时间为48个月（5~78个月）。第一次随机化后的观察指标是CR，第二次随机化后的主要观察指标是DFS，次要观察指标是OS和3级以上的不良反应。DFS定义为从CR的第一天开始到白血病复发或各种原因死亡，OS定义为从白血病诊断的第一天到各种原因死亡。Kaplan-Meier法估计DFS和OS，单因素预后分析采用Log-rank方法，多因素预后分析采用Cox回归模型。

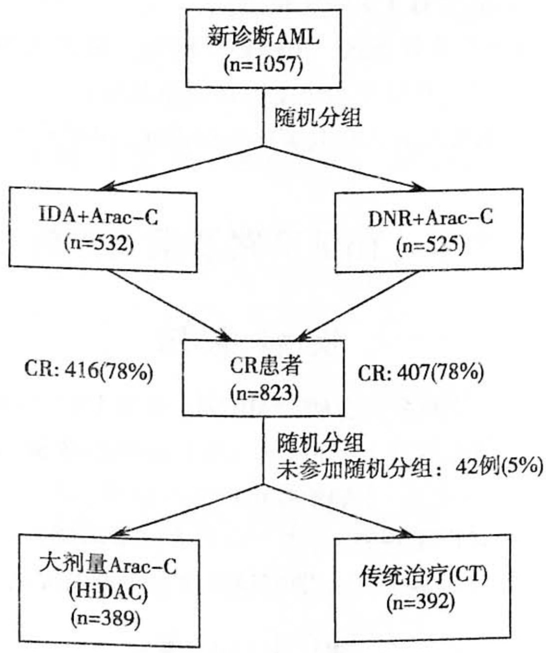


图11-6 CONSORT流程图

5. 研究结果 HiDAC组和CT组的5年DFS分别为43%和39%（$P=0.724$），5年OS分别为58%和56%（$P=0.954$）（见彩图11-7）。在染色体预后良好组，HiDAC组和CT组的5年DFS分别为57%和39%（$P=0.050$），5年OS分别为75%和66%（$P=0.174$）（见彩图11-8）。在染色体预后中等和不良组中，两组的DFS和OS无统计学差异。

应用Cox回归模型进行多因素预后分析，对于DFS而言，发病时白细胞$\geq 20\times 10^9$/L、诱导疗程数2次、年龄>50岁为独立的预后因素，校正了这些因素后缓解后的强化方案（应用传统化

疗）并非独立的预后因素。对于 OS 而言，年龄 >50 岁、诱导疗程数 2 次、发病白细胞≥20×10^9/L、原始细胞过氧化物酶（MPO）阳性率 <50% 为独立的预后因素，而校正了这些因素后缓解后的强化方案不是独立的预后因素（表 11-6）。

表 11-6　应用 COX 回归进行多因素预后分析

变量	分类	*HR*	*P* 值
DFS			
发病时 WBC 计数	≥20×10^9/L	1.49	<0.0001
诱导缓解次数	2 个疗程	1.50	0.0006
年龄	>50 岁	1.33	0.0028
强化方案	传统化疗	1.04	0.7128
OS			
年龄	>50 岁	2.00	<0.0001
诱导缓解次数	2 个疗程	1.58	0.0033
发病时 WBC 计数	≥20×10^9/L	1.41	0.0070
MPO 阳性的原始细胞比例	<50%	1.42	0.0149
强化方案	传统化疗	0.96	0.7768

DFS：无病生存；WBC：白细胞；OS：总体生存；MPO：过氧化物酶

在 HiDAC 组，白细胞减少更明显，持续时间更长，感染率更高（表 11-7）。

表 11-7　两组不良反应发生情况比较

	大剂量 Ara-C 组	传统化疗组	*P* 值
第一次化疗后最低白细胞（$\times10^9$/L）	0.17	0.40	<0.0001
第一次化疗后白细胞 <1.0×10^9/L 持续时间（天）	13（0~40）	12（0~36）	0.0005
感染发生率（%）	20.9	14.5	<0.001
出血发生率（%）	0.8	0.7	0.601
30 天死亡率（%）	0.9	0.6	0.389

6. 结论　完全缓解后，传统化疗与大剂量 Ara-C 治疗的疗效相当。在染色体预后良好的患者中 HiDAC 的 DFS 优于传统化疗。

（二）文献评价

1. 观察预后的研究对象是否都处于同一起始队列　是。急性白血病均为原发性、新诊断的病例，排除了继发性白血病和复发难治的病例。该文对 DFS 和 OS 有明确的定义，DFS 定义为从 CR 的第一天开始到白血病复发或各种原因死亡，OS 定义为从白血病诊断的第一天到各种原因死亡。

2. 研究的对象是否能代表被研究疾病的目标人群　是。该研究为多中心研究，研究对象来自 129 个研究单位，1000 多例病例，样本量大，代表性好。诊断标准明确，在各个研究中心诊断后，有一个专门的专家小组复核每一例纳入病例的血片和骨髓片，确保诊断的准确性。研究设计时计算了样本量，使把握度达到 80% 以上。

3. 随访时间是否足够？随访是否完整　是。该研究中位随访时间为 48 个月（5~78 个月），从图 11-7 可以估计中位 DFS 约为 20 个月，但尚不能得出中位 OS 时间，对于主要观察指标 DFS 而言随访时间足够，但对于次要观察指标 OS 而言，随访时间尚不够。该文没有直接报道失访率，根据文章描述，所有患者均有完整的随访结果。

4. 判断结局有无客观标准，是否采用了盲法　是。观察结局为 CR、DFS、OS 和不良反应，均有明确的定义。因为观察指标为客观的指标，不容易产生测量性偏倚，允许不采用盲法。

笔记

5. 是否对影响预后研究的重要因素进行了统计学的校正 是。应用了 Cox 回归模型进行了多因素的分析，对影响预后的主要因素进行了校正。

6. 预后研究的结果报告是否完整 是。报道了 CR 率、复发率、DFS、OS，采用 Kaplan-Meier 生存曲线形式展示了结果，也报道了 Cox 回归分析后的 *HR* 值和 *P* 值。没有报道具体的中位生存时间，但读者可以根据生存曲线进行估计。

7. 研究结果的精确性如何 不能确定。没有报道 95% *CI*，特别是表 10-6 中 *HR* 应该同时写出 *HR* 的 95% *CI*。

8. 我们自己的病人是否与文献报道的病人相似或非常不同 该研究为多中心、大样本量研究，亚洲日本病人，与我们中国病人应该相似，不会有很大的人种差异。

9. 研究结果是否有助于治疗方案的制订和是否有助于对患者及其亲属作出解释 是。该结果对临床医生提供了治疗决策，有助于对患者进行解释。

总体而言，该研究结论可信，有临床应用价值，可以指导患者的治疗。

四、实践决策

我们的 AML 患者染色体正常，属于染色体预后中等组，应用 HiDAC 和传统化疗可以取得相类似的长期预后，所以两种治疗方案均可以选择。与患者及其家属充分沟通后，患者采用 HiDAC 进行缓解后的强化治疗。但应用 1 个疗程后白细胞达到粒细胞缺乏、严重感染，以后换用了传统的化疗，直至缓解后强化治疗结束，患者持续缓解之中。

五、小　结

循证医学证据的最高级别是 meta 分析和 RCT，其次是前瞻性队列研究，该研究采用 RCT 的研究方法进行随机分组，保证了两组的可比性，再结合前瞻性队列的方法进行长期随访，随访率高，得出比较可靠的结论，指导临床决策。患者应用 HiDAC 后出现严重感染，因为有循证医学证据支持传统的常规化疗也可以取得相似的疗效，所以我们敢于换用常规化疗，事实证明患者取得了很好的效果。

（王小钦　林果为）

笔记

第十二章 临床实践指南的制定与评价

20 世纪 90 年代初期，美国医学研究所正式提出了临床实践指南(clinical practice guideline, CPG)的概念，经过 20 多年的发展，高质量的 CPG 为临床医生从事预防、诊断、治疗、康复、保健和管理工作带来许多益处，已成为国际上用来规范医疗行为、提高卫生服务质量、合理控制医疗费用等行之有效的方法。本章将介绍临床实践指南的基本概念及作用，查找指南的主要网络资源，并重点讲述临床实践指南的制作方法与流程，以及如何科学评价 CPG，最后介绍临床实践中正确应用 CPG 的原则和方法。

第一节 概 述

一、临床实践指南的概念和发展

临床实践指南，即根据特定的临床情况，系统制定的旨在帮助医生和病人做出恰当处理的指导意见。“特定的临床情况”主要指的是常见而重大的临床问题，如“高血压的防治”、“糖尿病防治”、“冠心病诊疗”、“脑卒中预防”等；“系统制定”即依据科学的方法规范化地制定指南的过程。CPG 的产生和发展主要与下列三个因素有关：

(一) 临床实践的巨大差异

自 20 世纪 80 年代以来，人们很快就注意到在临床实践过程中诊治上的差异现象。很多研究均发现对于相同一个临床问题，在不同国家或同一个国家的不同地区甚至在同一个地区内的不同医疗机构之间其处理方法也各式各样，差异性极大。例如一项关于中国和英国对急性缺血性脑卒中治疗方法的对比研究(表 12-1)，发现 19%~69% 的中国医师常规使用 7 种治疗方法，而英国医师在除使用阿司匹林外的其余 6 种疗法上的使用比例均不超过 1%。在美国，对于非瓣膜性心房颤动患者，南部与中西部的医生使用华法林预防脑卒中的比率差异达 4 倍之多；在同一个州内，一个社区的儿童扁桃体切除率是 8%，而另一个社区的切除率则高达 70%。临床实践的巨大差异已经超过了临床、人口学及地域上的差异所能解释的范围，令人们对这些差异现象的合理性及使用这些治疗措施的科学性产生怀疑。基于证据的 CPG 则可缩小这些差异，从而规范医疗行为，使患者得到应有且合理的医疗服务。

表 12-1 急性缺血性脑卒中治疗实践的中英对比

治疗方法	中国医师 (%)*	英国医师 (%)†	治疗方法	中国医师 (%)*	英国医师 (%)†
甘油 / 甘露醇	69	1	低分子右旋糖酐	44	0
中药	66	0	蛇毒	30	0
阿司匹林	54	39	激素	19	<1
钙拮抗剂	53	<1			

注：资料来自 Chen ZM, et al. Hospital management of acute ischemic stroke in China. J Stroke Cerebrovasc Dis, 1997, 6: 361-367. *回答常规使用各种治疗方法的中国医师的百分比；†回答常规使用各种治疗方法的英国医师的百分比

(二) 医疗措施的不当使用

有研究表明在日常临床诊疗实践中，大约 1/4 至 1/3 的医疗措施使用不当，存在着滥用(overuse)、误用(misuse)或使用不足(underuse)等问题。例如，抗生素对病毒所致的普通感冒几

笔记

乎没有益处，抗生素的使用不仅增加了患者的经济负担，而且长期使用还会增加副作用和导致耐药等问题。而美国的一项研究却显示，约半数的普通感冒和 2/3 的急性支气管炎患者接受了抗生素治疗，此为滥用。腹泻是临床常见的症状，在发病初期，腹泻能将肠道内的致病菌或其产生的毒素等有害物质排出，对于细菌感染引起的这类腹泻，止泻治疗不仅不能治愈疾病，反而会增加致病菌及毒素对机体的损害，但大多数患者常自行服用蒙脱石散、地芬诺酯等止泻药物，实为误用，有害无益。中国缺血性脑卒中和短暂性脑缺血发作的二级预防指南（2010 年）明确指出糖尿病患者且合并高血压时，降血压药物以血管紧张素转换酶抑制剂、血管紧张素Ⅱ受体拮抗剂在降低心脑血管事件方面获益明显。而我国的调查研究则显示，这类药物实际的服药率不到 50%，此为使用不足。

（三）医疗费用的日益增长

目前，全世界面临的难题是有限的医疗资源与日益增长的医疗保健需求之间的矛盾。2007 年我国卫生总费用达 11 289.5 亿元；2008 年为 14 535.4 亿元；而 2009 年达到 17 541.9 亿元，占 GDP 百分比为 5.15%。2010 年医院门诊病人和住院病人的人均医药费用，按当年价格比 2009 年分别上涨了 9.7% 和 9.0%。即使如美国这样的发达国家，庞大的医疗补助同样也让政府不堪重负。因此，明智地使用有限的医疗资源已经成为国际共识。根据科学的研究证据如成本 - 效益（cost-benefit）分析、成本 - 效果（cost-effectiveness）分析、成本 - 效用（cost-utility）分析、成本最小化分析（cost minimization analysis）等，制定一套规范化诊疗措施或临床实践指南，对于制定医疗费用补偿政策、合理高效地使用有限的医疗资源具有重大意义。

自 20 世纪 80 年代以来，诸如上述问题使各国政府和医疗保险机构感到了不堪承受的压力，因此认为 CPG 或许可以减少不恰当的医疗行为，改善患者预后，提高和保障医疗服务。CPG 的开发与应用自此被提上工作日程，并在许多国家得到迅速的发展。美国、加拿大、澳大利亚、中国香港特别行政区、新加坡及欧洲诸国均不同程度地建立了指南开发的相关机构，发展了诸多的临床指南，为解决卫生保健难题作出了巨大的贡献。遗憾的是我国的临床实践指南的开发还在起步阶段，目前尚无规范、统一和符合循证医学原则的指南开发专门机构及相关程序。

二、临床实践指南的作用

一个好的循证 CPG 需具备真实性、可靠性、可重复性、临床实（适）用性和明确的目标性等特征，一般是由各学科共同参与制作的、对当前最佳临床证据的评价与集成。临床实践指南的作用可归纳为以下六点：①提高医疗质量、改善临床结局；②概述研究发现的结果，并使临床决策透明化；③减少不同医疗机构和不同医师间医疗实践中的不恰当差异；④促进卫生资源的合理高效利用，从而减少患者的医疗费用；⑤区分医学研究结果的优先次序，作为医疗保险的凭证；⑥有助于医务人员的终身继续教育。

2011 年哈佛大学研究人员在《医院医学杂志》（*Journal of hospital Medicine*）上发表了一篇名为 “Use of UpToDate and Outcomes in US Hospitals” 的文献。它对比了 1017 家应用图文并茂的综合性循证指南 UpToDate 的医院和 2305 家未采用 UpToDate 的医院的医疗保险受益人数据，同时评估了这些医院的医疗质量和效率。研究结果表明，应用 UpToDate 的 1017 家医院每年可节省约 372 500 住院日，在三年考察期中累计挽救了约 11 500 例患者生命。根据全美医院质量联盟（Hospital Quality Alliance）所制定的指标来评定，发现应用 UpToDate 的医院提供了更佳的服务质量。研究提示：医生使用 UpToDate 系统，可帮助他们做出正确的即时诊断和治疗决策，可使医院患者驻留时间缩短、死亡减少且医疗服务质量得到提高。

三、指南与临床证据的关系

笔记

临床证据指的是循证医学中与临床实践密切相关的研究结果，这里的研究主要是针对人体

的，包括探讨诊断性试验准确性和精确性的证据、评估临床诊治安全性和有效性的证据以及研究疾病预后和预后因素的结果的证据等。因为各种类型临床证据的来源、科学性和可靠性是不同的，所以其论证强度也是不一样的。

这里介绍的是牛津循证医学中心（Oxford Centre for Evidence-based Medicine，OCEBM）的临床研究证据分级，亦属于新5级的证据分级法（表12-2）。

表12-2　牛津循证医学中心临床研究证据分级

推荐等级	证据分级	要求
A	1a	同质性的RCT的系统评价
	1b	可信区间窄的单个RCT
	1c	观察结果"全或无"
B	2a	同质性的队列研究的系统评价
	2b	单个队列研究，包括低质量的RCT，如随访率<80%
	3a	同质性病例对照研究的系统综述
	3b	单个病例对照研究
C	4	病例报告、低质量队列研究及病例对照研究
D	5	缺乏严格评价的专家意见，或基于病理生理和基础研究证据

原始研究证据和系统评价是客观地提供研究结果和对结果的解释，对临床实践有重要的参考和指导价值。循证医学提倡在临床实践中，尽可能使用当前可得到的最好证据、结合临床经验和患者的意愿进行诊疗方案的选择。加拿大McMaster大学临床流行病学与生物统计学Haynes教授提出了支持循证卫生决策的循证医学证据结构的"6S"模型。

"6S"模型即将原始研究（如PubMed数据库的原始研究）作为最底层，原始研究摘要作为第二层，综述（即系统评价，如Cochrane系统评价）作为第三层，然后是系统评价摘要（如循证期刊对原始研究和综述的简要描述），接下来是循证教科书中综述或研究总结、临床实践指南（summaries）。最后是系统（systems），系统通常是整合有CPG的计算机决策支持系统（CDSS），可根据个体患者的特征（如电子病历）链接相关证据，置于模型的最顶层。系统可提醒或告知医护人员治疗的决策。遗憾的是目前这种计算机决策支持系统寥寥无几，且未能做到及时更新。但可以预见这将是CPG网络资源利用研究的一个重要发展方向。

"6S"模型充分体现了临床实践指南在循证医学证据体系中的地位及其重要的临床指导意义。同样指南也是循证医学资源的一部分，它与原始研究证据和系统评价的区别在于：指南是针对具体临床问题，分析评价证据后提出具体的推荐意见，是连接证据和临床实践的桥梁和纽带。

四、循证临床实践指南的网络资源

21世纪的循证医学资源增长迅猛，临床实践指南（CPG）作为循证医学资源的特殊成员，其网络资源也日益增多。尽管医学界对临床实践指南网络资源的开发和利用现状还不尽如人意，但循证医学资源的有效利用和开发还是值得期待的。在此，我们将介绍主要的临床实践指南网络资源的分布及检索方法。

（一）原始研究证据数据库

美国国家图书馆开发研制的PubMed是循证医学原始研究证据数据库的代表，也是查找临床实践指南的重要资源。它收录了世界各国制作的指南，可使用"limit"辅助检索功能，只要在"Type of article"的下拉菜单中选中"practice guideline"即可检索临床实践指南。

中国知识基础设施数据库（China National Knowledge Infrastructure，CNKI）（http://www.cnki.net）含1100余种中文医药卫生类型期刊索引、引文索引及全文和报刊全文，更新较及时，可检索中

笔记

华医学会发布的各种临床实践指南。

（二）循证临床指南资源库

UpToDate（http://www.uptodate.com）创建于1992年，由美国Dr.Burton D.Rose和Dr.Joseph Rush开发，隶属于Wolters Kluwer health公司，该公司为医疗保健行业提供信息、商业智能和即时解决方案，是Wolters Kluwer旗下的公司。而Wolters Kluwer则是一家全球领先的信息服务公司。UpToDate是图文并茂的综合性循证指南、患者手册及药物间交互作用查询系统，给医师、药师提供即时、实证的临床医药信息，并可以快速解答临床专业人员提出的临床问题。文献中附有图片，包括图表、X线片、相片、影像档等，及MEDLINE的引用文献摘要。UpToDate主题包括内科学、妇产科学等二十个类别，每个主题之下尚有更专精的类别，其内容系由网罗众多相关的文献而成，可以快速地获得临床上最前沿医学问题的答案。

EBM Guideline（ebmg.wiley.com）由芬兰医学协会开发和维护，隶属于Wiley Blackwell公司，该公司成立于2007年，是世界领先的学协会出版商，出版大约1350种同行评审学术期刊及涵盖面广泛的具有全球影响力的书籍。涵盖学科领域包括科学、技术、医学、社会科学及人文，为广大科研教学人员提供最新的高质量的信息。EBM Guidelines包含有综合性循证指南、证据总结、医学图片、影音，收录涵盖实证数据的基础医疗临床指引，提供最新的临床医学发展及实证资料，内容全面涵盖了全科医师经常遇到的临床问题，并提供诊断结果及治疗方法（包括相应的推荐意见及强度）。可链接至高质量的照片影像及实证文献。

（三）临床实践指南网站

世界各国都有相应的网站专门提供临床实践指南检索。其中比较大型且权威的有美国国立指南数据库和加拿大医学会临床实践指南网站等。

1. 美国国立指南数据库（National Guideline Clearinghouse，NGC）（http://www.guideline.gov）由美国卫生健康研究与质量机构（AHRQ）、美国医学会（AMA）和美国卫生健康计划协会（AAHP）于1998年联合制作。它是一个提供CPG和相关证据的且功能完善的免费数据库。NGC提供直接检索和浏览两条检索途径，并可对收集的指南进行比较。

2. 加拿大医学会临床实践指南网站（Canadian Medical Association Clinical Practice Guidelines），其数据库（http://www.cmaj.ca/misc/service/guidelines.html）由加拿大国家、州或地区医学卫生组织、专业协会、政府机构和专家小组于1995年共同主办。指南由加拿大各地和各机构团体提供。网站的栏目有：开发者名单、最新内容、热门话题、新闻、方法与资源、其他CPG网站、常见问题解答、搜索帮助、用户反馈及提交指南说明等。网站上还链接有加拿大医学会制作的《临床实践指南手册》。网站中一半以上的指南有全文。

3. 苏格兰校际指南协作网（Scottish Intercollegiate Guidelines Network，SIGN）（http://www.sign.ac.uk/guidelines）建于1993年，为苏格兰国家卫生服务开发临床实践指南。重点关注癌症、心血管疾病和心理卫生等领域。网站的栏目有：指南（分为按主题排列的指南和按索取号排列的指南）、指南选题提示或范围、当前指南项目组正在进行的工作、指南开发的方法学等。此外，该网站还链接有指南制作的支持材料、简介、用户申明及版权细节等内容。网站提供指南全文。

4. 新西兰指南研究组（The New Zealand Guideline Group，NZGG）（http://www.nzgg.org.nz）由新西兰卫生部在1996年组建，主要目的是为了制定和实施循证临床实践指南。内容主要包括：用于实践的证据、消费者资源、新西兰循证健康公告等。其将指南分为四种类型：基层医疗服务管理指南（guidelines for primary care management）、病人转诊和管理指南（guidelines for patient referral and management）、第一专科评估准入标准指南（guideline for access criteria for first special list assessment）和临床优先评估标准指南（guideline for criteria for clinical priority assessment）。此外，该网站还链接一系列与临床指南的开发和评价有关的网站，如：证据源、Cochrane协作组、循证方法及根据、临床指南等网站。

笔记

（四）循证医学期刊

ACP Journal Club（http://www.acpjc.org/）由美国内科医师协会于1991年创办。网上可免费获取全文。在ACP Journal Club网站上搜索有关专题的临床实践指南，只需在检索框中输入某个专题，在检索框下的“article type”中选择“clinical prediction guide”即可。

（五）循证医学多元搜索引擎

1. SumSearch（http://sumsearch.uthscsa.edu）由美国Texas大学的卫生科学中心建立并维护，同步检索Cochrane图书馆、免费PubMed、NGC等。最大优点是能帮助临床医师快速获得所需证据，对临床实践很有帮助。可同时检索系统评价和临床实践指南，可针对治疗、诊断、病因、预后等方面进行限定检索。

2. TRIP Database（http://www.tripdatabase.com）1997年建立，可同步检索数十种证据源，包括循证摘要、循证指南、二次研究、患者手册等。该数据库提供分类浏览和关键词检索两种检索途径。其Clinical Areas下列有13个分类专题目录，如Cancer、cardiovascular等。用户可以选择点击浏览相关内容。关键词检索分为简单检索和高级检索两种方式，检索方便，结果层次清晰。

3. Doctor Desk（http://drsdesk.sghms.ac.uk）由英国国家保健服务（系统）卫生保健电子图书馆（NeLH）建立，网站立足于简单易用，主页上有EBM Search检索框，输入检索词即开始检索循证医学方面的指南、系统评价或研究论文。指南来自SIGN、NICE、PRODIGY和NeLH。检索功能较为完善，检索词之间可以用“AND”、“OR”、“NOT”逻辑运算符。

第二节 临床实践指南制定的方法与流程

一、临床实践指南的制定方法

一般分为两类：

（一）专家共识制定法（consensus guideline development）

分为非正式和正式的专家共识制定法。非正式的专家共识制定法是由一组专家开会讨论，将一次或多次开会讨论后达成的共识形成推荐意见，再由专业学会或政府机构进行发布。这种共识只包括推荐意见而缺乏形成推荐意见的证据基础和制定共识的背景及方法，易受参会人员的专业、优势、性格、组织和政治因素等影响，可靠性和质量较差。

正式的专家共识制定法，首先是应邀专家召开预备会，就某一治疗措施专家组提供相关研究证据的综述并形成可能的适应证清单。然后在第一次正式会议时专家组成员对每个适应证进行评分以评价其适用性，评分通常使用手填量表的方式，量表分为1分到9分，1分为完全不适用，9分为特别适用，5分为可用或不用。再次开会时专家们将小组集体评分的情况与自己的评分相比较，讨论不一致的原因，然后再次重复评分，在会议讨论的基础上修改评分。最后的评分反映了专家组成员的一致性程度。这种制定法虽然考虑了研究证据，但仍然没有将推荐意见与相关证据的质量明确联系在一起，在实践中临床医生难以恰当选用。

（二）特殊指南制定法

分为明晰指南制定法（explicit guideline development）和循证实践指南制定法（evidence-based guideline development）。

明晰指南制定法是由指南制定者对治疗措施的益处、危害及费用作出说明，并明确地估计出现每种结局的概率。估计值的来源也一定会在文件中具体写明。因为这种分析方法非常复杂，对指南制定者来说太烦琐，目前使用此方法已不多，在实践应用中也不适用。

循证实践指南制定法（evidence-based guideline development），是目前国际推崇的指南制定方法，即将推荐意见与证据质量明确地联系在一起，依据现有证据来确定推荐意见的强度。这也

笔记

是循证临床指南的明显特征，是保证指南的科学、公正和权威的方法。以下将重点介绍此方法的制定流程。

二、循证实践临床指南制定流程及步骤

苏格兰校际指南协作网（Scottish Intercollegiate Guidelines Network，SIGN）是目前推荐的循证实践指南制定的代表机构。

制定流程为：由指南开发组织确定指南题目和意义——成立指南制定小组——检索并系统评价文献（证据）——形成指南推荐草案——咨询并进行同行评审——发布指南——临床应用指南——再评估。其制定指南的主要步骤如下：

（一）确定指南目的和意义

指南开发组织机构首先确定指南拟解决问题的重要性（如发病率、结局效果、经济费用）及制定指南的必要性、目的和使用范围。

（二）成立指南制定小组

指南制定小组人员一般要求来自不同地区的多个学科，一般为 15~20 人，包括：医生，专业编辑人员，信息专家，IT 专业人士及市场和培训人员。并确定制定指南的规范程序。指南制定人员的素质应具备四个核心技能：临床专业技能；卫生保健的实践经验；专业知识技能；严格的评估技能。

（三）检索文献（证据）

全面收集世界范围内的相关研究资料。一般是先检索已有的临床指南，再检索最新的系统评价证据，最后是原始研究证据。信息专家负责审查检索结果的质量，经审查合格的文献再送交制定指南的专家小组。

（四）评价文献（证据）

指南制定专家对证据提出推荐意见并对推荐强度进行标注。指南制定小组首先制定一套明确的文献纳入和排除标准，然后对最终纳入的文献，采用一套严格的文献评价系统来评价文献的质量，同时对文献进行证据分级。

首先，对证据内涵和质量的评价，一般包括如下方面：①证据的一致性：包括各个文献的总体一致性；各文献入选人群特征（如年龄、性别、宗教信仰等）的一致性；研究内容的一致性。②外部真实性：研究结果是否与实际应用时的结果一致或者相反。③针对性：证据是否直接针对指南的目标人群。④证据容量：即每个研究人群的样本数量和所有相关研究的样本数量。

其次，对证据的解释，一般包括如下方面：①病人意愿：权衡指南应用后的利弊；对患者预后的改善程度。②临床实际：证据是否与当地现有的医疗实际情况有极大的差距。③资源分配：证据是否会导致大规模的资源重新分配。

在文献评价中，每一篇文献至少应有 2 名指南制定小组成员进行评价，如果评价存在分歧，则由第三者仲裁解决。最后，经过严格的证据评价达成共识，并且根据支持证据的强度决定推荐的等级，从而制定出指南初稿。

（五）形成并评审指南草案

指南草案形成后，将提交给指南制定小组外的同行专家团进行评审。这个评审专家团将对指南初稿做出评价并对初稿中存在的问题向指南小组提出疑问及建议。指南制定小组根据建议进一步修订指南。修订版再次送同行专家团进行评价。最后，由 SIGN 编辑组对指南进行审查并做出最终评价。

（六）发布指南及评估

当指南正式成文后，可制成各种版本发布，供不同的对象使用。除了全文外，可以是摘要性的结论性建议，也可以是针对患者的教育手册。当指南发布后，制定专家小组将负责推广和

笔记

指导指南的使用，同时监督和评价指南的执行情况，收集反馈信息。

（七）指南更新

每个指南发布2年后需要进行再评估。将根据该领域的最新进展以及后效评估的反馈意见，决定是否更新指南。

指南的制定是一个十分复杂的过程。在这个过程中，还需要考虑患者和护理者的参与，并收集他们的建议，以确保他们的意愿在指南中得到体现。由于制定每一个指南需耗费大量的金钱和时间，指南文件形成后应注意保存指南制定过程中形成的文件，以便后期指导或评估时使用。例如：制定指南的原始提议、制定指南的缘由、指南制定中的关键问题及检索策略和数据库、文献的纳入及排除标准、回答所有关键问题的证据总结表、指南小组对证据质量和相关建议分级、会议总结及同行评议和回复记录。

其中，循证实践临床指南具有科学客观的证据依据，令人信服；同时又标注了推荐意见的强度，便于使用者根据其强度是否采用其推荐意见。虽然制定临床实践指南要求使用系统评价的证据作为基础，但当一些常见和重要的临床问题还缺乏充分的A级证据时，指南可以基于稍差的证据，但必须取得小组成员的共识。这样的指南既贴近现实，又可被接受。

总而言之，一个好的指南应具备两个特点：一是对证据进行综合及概括，得到某种干预措施对典型患者的平均效果证据。二是说明如何将这一证据用于形成推荐意见。针对推荐意见需要具体说明干预措施的利弊、局限性、最适宜的患者及人群，以及与成本和卫生保健有关的其他因素。

第三节　临床实践指南的评价

对同一种疾病，不同的国家或学术组织可能制定出不同的指南，指南的质量也可参差不齐，甚至某些建议互相矛盾，这些都将给临床决策带来极大困扰。因此，在使用指南之前，临床医生应有评价和鉴别其质量高低的能力，判断该指南是否值得推荐使用，或者从众多的指南中挑选出质量最好的应用于临床实践。以我国2011年"循证医学与临床实践指南制定、评价和解读方法学研讨会"上的数据为例，我国2010年共有指南75篇，经质量评价后，发现完全符合评审标准的只有2篇、基本规范的有7篇、无法评价的有66篇。

一、临床实践指南的评价

首先，评价一个临床实践指南（CPG）的好坏，应强调以证据为基础，即真实性。好指南必须使用循证医学的原则和方法，根据证据的可信程度对每一条建议进行分级。世界著名临床流行病学家、循证医学的奠基人之一David Sackett指出，确定指南的质量好坏主要依据两个方面：①指南是否收集了所有最新（12个月）的有关证据，并对其进行了分析评价，是否对其真实性进行了分级？②指南是否对每一条推荐建议标注了其依据的证据级别和相关文献出处？因此，对指南的质量评价实际上主要集中在证据的收集、评价和合成，以及如何将推荐意见与相关证据进行集成的过程。

其次，当指南的质量真实性得到肯定后，就应对其重要性进行评价。即该指南是否回答了临床需要解决的重要问题？这些问题必须是临床医生在工作中面对的问题。当然，由于临床所面临的问题相当复杂，一个指南不可能囊括所有的临床问题，只能是绝大多数重要问题。

最后，应用指南前还要评估其适用性。一个指南的成功实施依赖于四个因素（4B）：

1. 疾病负担（burden of disease）　如该病在本地区的发病率很低带来的疾病负担低，因而无需使用参考指南？我的病人是否不可能出现指南中所描述的结果？如果是这样，则应用指南不仅浪费时间金钱，还可能造成不必要的伤害。反之，则应积极使用指南。具体到个体病人时，

笔记

还应综合考虑:①我的病人与研究证据中的病人的情况差别大吗?②在现有的环境条件下这种治疗方案可以实施吗?③指南中推荐的治疗方案有哪些潜在的利弊?

2. 患者对治疗价值的看法(beliefs),对治疗方案的利弊评价是否有别于指南中推荐的内容?

3. 执行该指南所推荐的措施所需成本(bargain)有多大,机会成本如何?其他干预措施能否有更大的获益?

4. 对患者实施该指南是否存在不可克服的障碍(barriers)?这些困难包括地域性的(如该地区根本不能实施此治疗方法)、传统性的(如长期惯用另一种治疗方法)、权威性的(教授怎么说就得怎么做)、法律性的(医生惧怕因为舍弃了常用但效果不明显的疗法会遭到起诉)、行为性的(医生无能为力或病人不能服药)。若这些障碍明显,则无法实施指南。

总之,指南的推荐意见是原则性的,应在一般原则指导下,实施个体化诊治,同时还应结合患者(或其家属)的需求及价值取向。

二、临床实践指南的评价工具

许多国家和学术团体都制定了CPG的评价工具以便科学客观地评价CPG。目前,得到公认的指南评价工具有两个:美国COGS(the conference on guideline standardization)评价标准和欧洲AGREE(appraisal of guidelines research and evaluation)量表。

(一)美国COGS评价标准

美国医学研究所(Institute of Medicine, IOM)早在1990年发表了第一个针对CPG的评价工具,共8条评价标准,分别是效度、信度、临床实用性、临床灵活性、透明度、多学科联合开发、指南定期评价、指南开发。2002年美国COGS会议确定了18条评价指南的标准。表12-3总结了美国COGS会议制定的18条评价标准。

表12-3 GCP报告的COGS清单

条目	描述
1. 概述	提供结构性摘要包括发布日期,指南情况(原版、修订、更新)、印刷和电子版
2. 重点	描述指南涉及的主要疾病和干预措施,指出可替代的预防、诊断和干预措施
3. 目标	指南希望达到的目标,和达到这一目标的理由
4. 使用者/背景	描述指南的使用者和指南应用的背景
5. 目标人群	适合指南推荐的患者人群并列出排除标准
6. 制定者	区别指南制定机构与指南制定个人的潜在利益冲突
7. 资金来源或赞助人	确定资金来源或赞助人在指南制定和发布中的作用,说明潜在的利益冲突
8. 收集证据的方法	描述文献的检索方法,包括日期、数据库和检索标准
9. 建议分级标准	描述证据质量的分级标准和推荐强度的分级系统。推荐强度与推荐的重要性相关,并基于证据的质量和预期获益或损害的大小
10. 综合证据的方法	描述证据如何被综合为推荐意见,如证据表格、meta分析、决策分析
11. 发布前评审	描述指南发布前是如何进行评审的
12. 更新计划	陈述是否有指南更新计划,并标注本指南的有效期
13. 定义	描述不常用的术语,并严格纠正易被指南误解之处
14. 建议与基本原则	准确地陈述指南的作用和执行指南的特殊情况,通过描述证据与推荐之间的联系来判断每一项推荐,根据第九条来显示证据的质量和推荐的强度
15. 潜在利弊	应用指南预期的获益和潜在的风险
16. 患者意愿	当指南涉及个人选择或价值时应考虑患者的意愿
17. 法则	必要时提供图表说明
18. 指南在执行中需要考虑的事项	描述指南推荐应用时的障碍,对任何辅助文件给予参考文献以便于应用

笔记

（二）欧洲AGREE量表

指南研究与评价工具AGREE(appraisal of guidelines research and evaluation)由13个欧洲国家的研究者制定。该评价工具在国际上具有较高的权威性，为目前国际指南质量评价的基础工具。可从http://www.agreecollaboration.org下载获得。

AGREE量表包括6个维度(指南的范围和目的、利益相关者的参与度、指南开发的严格性、指南的清晰性与明确性、指南的适用性和编辑工作的独立性)共计23个条目(表12-4)。

表12-4 AGREE评价工具

维度一：指南的范围和目的(涉及指南的总目的、具体的临床问题和适用的患者)
1. 明确说明指南的撰写目的
2. 明确阐述了指南所涵盖的临床问题
3. 明确阐述了指南所要应用的目标人群，应提供年龄范围、性别、临床类型及共病
维度二：利益相关者的参与度
4. 指南制定小组的成员包括所有相关方面的专家
5. 指南考虑了目标患者的观点和选择
6. 指南的适用对象已经清楚说明
7. 指南发表前已经在目标使用者中进行了试行
维度三：指南开发的严格性(涉及收集和综合证据的过程，制定和更新推荐建议的步骤方法)
8. 采用严格系统的方法检索证据，应提供检索证据的详细策略，包括使用的检索词、信息来源和文献涵盖的时间
9. 证据的纳入和排除标准有清楚的描述
10. 详细描述了形成推荐建议的方法
11. 形成推荐建议时充分考虑了干预措施的获益、风险和副作用
12. 推荐建议与支持证据之间有明确的联系
13. 指南发表前接受了制定小组以外专家的同行评议
14. 提供了更新指南的程序
维度四：指南的清晰性与明确性
15. 推荐建议明确而不含糊
16. 针对不同临床状况提供了不同的治疗选择
17. 关键的推荐建议能容易识别
18. 提供了指南方便应用的工具(如概要、教育工具、针对患者的宣传单等)
维度五：指南的适用性(指南应用时可能涉及的单位、操作和费用问题)
19. 讨论了应用该指南的过程中可能遇到的困难
20. 讨论了应用该指南的过程中可能的费用
21. 提供监督和审查指南执行情况的评价标准
维度六：编辑工作的独立性(推荐建议的独立性和对指南制定组各成员利益冲突的说明)
22. 指南编辑工作独立于其他经济实体，不受基金资助机构的影响
23. 指南制定小组成员的利益冲突或竞争性利益的声明

表中每个条目可按1~4分打分，完全符合条目要求的打4分，完全不符合的打1分，介于二者之间的根据测评人员的判断给2分或3分。在评价中，为了避免审查者主观性影响，建议由2~4名参评人员来评定，以增加指南评估的可靠性。同时在每项条目后都提供了补充说明的信息，有助于参评人员对条目所涉及问题和概念有更好的理解，评分正确合理。最终结果将根据公式(表12-5)对所有参评人员的评分进行综合，分别得到6个维度中每一个部分的标化百分比。

笔记

表 12-5 第一维度得分的计算方法

	条目 1	条目 2	条目 3	总分
评估员 1	2	3	3	8
评估员 2	3	3	4	10
评估员 3	2	4	3	9
评估员 4	2	3	4	9
总分	9	13	14	36

可能的最高分数 = 4(完全符合) × 3(项目) × 4(评估员) = 48 分
可能的最低分数 = 1(完全不符合) × 3(项目) × 4(评估员) = 12 分
该维度的标化百分比 =(实际总分 − 可能最低分数)/(可能最高分数 − 可能最低分数)
= (36 − 12)/(48 − 12) = 67%

根据 6 个维度标化百分比的比值大小判断该指南是否值得推荐应用，结果分 3 个等级：①强烈推荐：大多数条目为高分(3~4 分)，并且 6 个维度中大多数标化百分比 > 60%；②推荐：低分条目(1~2 分)与高分条目(3~4 分)数目大致相当，6 个维度中大多数的标化百分比介于 30%~60%；③不推荐：大多数条目为低分(1~2 分)，6 个维度中大多数的标化百分比 < 30%。应注意 6 个维度的得分是独立的，不宜合并成一个质量评价的总分值。

AGREE 量表的不足在于缺乏具体标准来评价指南的临床内容，缺乏对推荐意见合理性的审查，因此还需要专家评议和临床试用来补充评估。尽管如此，AGREE 量表仍是评价 CPG 的首选工具。AGREEⅡ为最新版本。

第四节 临床实践指南应用的原则和方法

一、临床实践指南的应用原则

临床实践指南是为临床医生处理临床问题制定的参考性文件，是推荐应用而非强制执行，应避免不分具体情况盲目地、教条地照搬使用。应用时注意考虑以下原则：

（一）个体化原则

在应用指南时，医生应充分考虑该患者临床特征是否与指南中目标人群一致。面对具体的个体患者，临床医生应该在指南的指导下，综合具体病情和多方面的因素个体化地选择治疗方案。临床医生应具备以下两方面素质才能保证指南的正确使用：第一，快速判断患者病情状况和建立诊断的能力；第二，具有对患者接受干预措施可能获得的利弊进行评估的能力。例如冠心病合并糖尿病患者根据指南应强烈推荐使用阿司匹林，但如患者患有严重的消化道溃疡病，阿司匹林需慎用甚至不用。

（二）适用性原则

自己患者的情况与指南的目标人群相似吗？如果相似，可以考虑应用指南推荐的干预措施。并结合本地区或医院目前的医疗条件，患者的经济状况，评估其对医疗费用的承受能力，以及当地医疗保健系统的覆盖支持能力等因素，评估该干预措施的可行性和费用 - 效益比。例如各国指南都推荐急性心肌梗死早期(3~12 小时内)行经皮冠状动脉介入(PCI)治疗，但在我国绝大多数基层医院无条件开展此项技术，多数心肌梗死患者也不能承受高昂费用，此时的诊治就只能先采取指南建议的药物治疗方案，待病情适宜条件下转上级医院治疗。

（三）患者价值取向原则

患者或其亲属的价值取向和意愿在临床决策中具有重要的作用。医生应事先与患者或家属沟通，了解他们期望的结局指标以及与指南中的结局指标是否一致，差距有多大。应充分尊

笔记

重患者或家属的选择。例如对 1 例晚期肺癌患者，应用体力状况（performance statue，PS）评分标准进行评估，得到评分为 3 分（PS 评分一般要求不大于 2 分才可能实施有效的抗肿瘤治疗），在这种情况下指南不支持使用细胞毒类药物，因为放化疗与姑息治疗对生存期没有改善，甚至严重副作用会影响患者的生存质量。但患者及家属坚持认为不积极治疗就等于放弃，最后还是尊重患者及家属的意愿，试行放化疗，并予以严密监测患者病情变化。

（四）时效性原则

随着医学的快速发展，基础和临床研究证据也在不断更新。过去认为有效的治疗手段可能被新的证据证明无效，而过去认为无效甚至禁忌的治疗手段可能被新的证据证明有效。例如既往认为慢性充血性心力衰竭是使用 β- 受体阻滞剂的禁忌证，但目前大型随机对照试验一致证实 β- 受体阻滞剂可以显著改善慢性心力衰竭患者的预后。因此，现有指南（Ⅰ类推荐，A 级证据）β- 受体阻滞剂治疗慢性充血性心力衰竭。

（五）后效评价原则

后效评价是指在患者接受根据 CPG 制定的诊治方案后，继续临床随访以评价患者病情的变化。后效评价在整个循证临床实践中具有重要作用，也可为指南的修订和更新提供临床资料。例如目前对于乙型肝炎的治疗，恩替卡韦疗效确定，已成为一线抗病毒用药，但当患者开始使用后，仍需定期复查肝功能及病毒 DNA 量，以评估药物的有效性，及时发现可能出现的耐药情况。

二、临床实践指南的应用技巧

在实际临床工作中，临床医生应用 CPG 应注意以下技巧：

1. 明确该指南制定的方法，一项真正的、以循证为基础的 CPG 较非循证 CPG 的可靠性更强。

2. 了解并分析指南中的证据水平与推荐建议强度，并明确它们之间的关系，以便判断推荐意见的可靠程度。

3. 依据推荐意见强度确定是否应用于临床。

4. 注意消除指南的实施障碍，避免指南使用不足。指南在实施过程中总会遇到一些障碍，如：①社会因素障碍，像某些新的治疗措施社保不予支付；②个人因素障碍，如医生过度自信，盲目使用指南；③环境因素障碍，被医药宣传信息误导，脱离指南、习惯性给予“常规治疗”等。

此外，在应用指南过程中，应充分重视个人沟通能力的重要性，加强与患者及家属的沟通交流，既可增加患者及家属对指南应用的理解，也是临床医生正确使用指南、进行个体化临床决策的基础。同时在选择指南时，应尽可能选择由本地区或本国制定的指南。但遗憾的是我国制定的指南大多为传统的专家共识，指南质量偏低。而在选用欧美国家指南或国际性指南时，应充分估计到本地人群的基线特征、医疗卫生资源的分布，这些都可能与指南存在差异，影响指南的具体应用。最后，应用指南时应充分体现循证医学的理念，即根据患者的具体临床情况，将当前所获最佳证据与临床技能和经验相结合，考虑成本 - 效益比及当地卫生资源的实际情况，并在充分尊重患者及其亲属的价值取向和意愿的基础上，做出最佳的、综合性的临床决策。

（廖晓阳　刘金来）

笔记

第十三章　循证医学的个体化实践

循证医学是把新近最佳证据用于临床适合的患者，并取得最佳临床诊治效果的过程。与传统医学的区别在于，循证医学采用科学的标准严格筛选、分析和评价拟用临床研究证据，推崇使用最佳证据，适时更新证据，并对临床诊治效果进行后效评价，进而发现新问题，开始新一轮的循证实践过程。

任何临床研究证据和成果都是源于临床对患者个体特征的观测或对试验性干预反应的总结，是从基于个体特征的感性认识升华为对群体的共性认识。证据或研究结果就群体而言是真实的，其推广和应用对改善和提高患者群体的疾病防治效果会有积极的作用。然而针对自己的患者个体，可能与证据来源群体存在特征上的差异，比如地域、种族、年龄分布、疾病状态、社会经济和环境背景等方面。能否将从群体获得的最佳证据或研究结果用于具体的个体患者，实现将患者个体化特征、意愿 / 价值观以及现有最佳证据的有机结合，是临床医师在实践循证医学时需要重视的问题，也是将最佳证据转化为最佳临床诊治效果的重要环节。

第一节　临床最佳证据的特性

实践循证医学的证据，应该是新近最佳证据（best evidence）。基于临床流行病学和循证医学的方法和原则，最佳证据应具备以下特征：

一、真　实　性

证据的真实性（validity）指研究结果与实际真实结果间的符合程度，即研究结果的准确性。由于研究设计、实施和评价过程有诸多影响因素，导致研究结果或多或少存在与实际结果之间的偏差，这就是为什么同样的研究设计会得出不同的结论的原因。为避免伪证据或夸大效果的研究结论对临床诊治的影响，对拟采用的证据首先必须评价其真实性。比如，研究报道某种药物治疗高血压效果显著，但对其分析评价发现，纳入的研究对象没有严格的诊断标准，设立的对照组与试验组基线存在明显的差异，临床医师应对此研究结论提出质疑，这样的研究结果究竟是药物真实的降压效果还是选择偏倚造成的，能否用于临床指导值得推敲。如以原始研究证据为例，基于影响研究结果的主要环节，从以下几方面进行真实性评价：①研究的设计方案是否合理，是否设立恰当的对照组；②研究对象的诊断标准及其纳入和排除标准是否明确；③组间的基线是否可比、干预措施和方法是否科学有效和安全；④终点指标是否确切、恰当；⑤干预措施实施及结果测量、分析和报告是否采用盲法；⑥研究结果有何偏倚存在及采取了哪些防止和处理方法；⑦患者的依从性如何，随访率多少；⑧资料收集、整理、统计分析方法是否合理等。根据不同的设计方案及可能产生的偏倚大小，可对拟采用证据的论证强度和可信度水平进行分级，在具体应用中优选高级别的证据。研究证据使用的推荐强度分为 5 级，即Ⅰ级、Ⅱ级、Ⅲ级、Ⅳ级和Ⅴ级，具体分级见表 13-1。

笔记

表 13-1　牛津证据分级与推荐强度（2001 年）

推荐强度	证据级别	治疗、预防、病因研究	预后研究	诊断性研究	经济学分析
Ⅰ级	Ⅰa	基于同质性随机对照试验的系统评价	基于同质性前瞻性队列研究的系统评价，或经验证的临床实践指南	基于同质性一流水平的诊断性研究的系统评价，或经验证的临床实践指南	基于同质性一流水平的经济学研究的系统评价
	Ⅰb	可信区间窄的随机对照试验	随访率≥80% 的前瞻性队列研究	纳入研究对象适当，且与金标准进行了同步独立盲法比较的诊断性研究	采用适当的成本计算，对所有经过严格验证的备选医疗方案的结局进行了比较分析，包括敏感性分析
	Ⅰc	观察结果为“全或无”（某干预措施推行前某病病死率为 100%，推行后低于 100%，或推行前某病患者存在死亡或治疗失败现象，推行后无死亡或治疗失败）	观察结果为“全或无”的病例系列研究	绝对的特异度高即阳性者可确诊，或绝对的灵敏度高即阴性者可排除	对干预措施分析后有明确结论：①成本低、其结果佳的程度；②成本高、其结果差的程度；③成本相同、其结果的好坏程度
Ⅱ级	Ⅱa	基于同质性队列研究的系统评价	同质性回顾性队列研究，或对照组未治疗的、基于随机对照试验的同质性系统评价	同质但水平低于Ⅰ级的诊断研究系统评价	同质但水平低于Ⅰ级的经济学研究系统评价
	Ⅱb	单个队列研究（包括低质量的随机对照试验，如随访率低于 80%）	回顾性队列研究，或随机对照试验中未作治疗的对照组患者随访结果，或未经验证的临床实践指南	同步作了金标准及诊断试验，并进行了独立盲法比较，但研究对象局限且不连贯，或未经验证的临床实践指南	采用适当的成本计算，对若干备选医疗方案的结局进行了比较分析，包括敏感性分析
	Ⅱc	结局性研究*	结局性研究*		
Ⅲ级	Ⅲa	基于同质性病例对照研究的系统评价	—	—	—
	Ⅲb	单个病例对照研究	—	纳入研究对象适当且与金标准进行了独立盲法比较或客观比较，但部分内容对象未接受金标准试验的诊断性研究	未作准确成本计算的经济学研究，但进行了敏感性分析
Ⅳ级		系列病例观察（包括低质量的队列研究和病例对照研究）	系列病例观察（包括低质量的预后队列研究）	未采用盲法或未客观独立地使用金标准试验的诊断性研究，划分真阳性和真阴性的参考标准不统一、或纳入研究对象不适当的诊断性研究	无敏感性分析的经济学研究
Ⅴ级		专家意见或基于生理、病理生理和基础研究类证据	专家意见或基于生理、病理生理和基础研究类证据	专家意见或基于生理、病理生理和基础研究类证据	专家意见或基于经济学理论类证据

注：* 结局性研究是指描述、解释、预测某些干预或危险因素对最终结局作用和影响的研究

笔记

二、重要性

研究结果的重要性(importance)指其临床价值和意义。重要性评价应建立在证据真实的基础上,否则不真实的证据就没有重要性可言。不同类型的研究证据,重要性有着不同的表达形式。

对于诊断性试验,重要性的指标反映在敏感度提高了多少,敏感度高的诊断试验表明漏诊率低,用于疾病的筛查效果好;另一个指标是特异度提高了多少,特异度高的诊断试验误诊率低,用于疾病的肯定诊断。了解诊断试验的意义,有助于自己在疾病的诊断中,选择恰当的诊断方法,从而提高临床诊断的准确度。

在治疗性研究的证据中,需要了解实施了干预措施后,能使疾病的治愈率提高多少或者病死率及病残率降低了多少,干预措施的安全性如何。在权衡利弊后,选择安全有效的措施。

在疾病预后的研究中,要分析不同预后因素(有害或有利因素)对疾病预后的影响及其程度。

对于不同的研究结果还应该考虑其成本 - 效果。比如诊断性试验的应用,使每增加一例确诊病例所增加的成本是多少,或治疗性研究中每防止一例不良结果或并发症所增加的成本等。

采用定性和量化的指标将有助于判断临床研究结果的重要程度。

三、适用性

任何最佳证据的应用和推广,都必须结合病人的实际情况、医疗条件、医务人员的知识技能水平、患者的接受程度以及社会经济状况的承受能力等,所以适用性(applicability)关系着证据应用和推广的可行性。

来源于不同国家和地区的临床研究证据,由于人种及生物学上的差异,可能导致不同反应的出现;社会文化背景的差异可能影响证据的接受和推广;经济发展水平对新技术和新药物的应用有着明显的促进或制约;医务人员的知识水平和医疗条件在不同国家和地区的差异也会影响证据的使用。对来源于不同国家和地区的临床研究结果,其引用要充分考虑上述的差异,一定要结合自己病人的特点,具体分析相关证据的应用价值及可行性,不能盲目照搬。

四、经济性

物美价廉在医疗保健领域同样适用,全球都面临着卫生资源有限而卫生保健需求无限这一矛盾。如何合理利用有限的卫生资源来满足人们对健康保健日益增长的需求,不仅是国家层面需要研究和解决的问题,同时也是临床医师在临床实践中需要面对的问题。就临床最佳证据而言,在真实性、重要性和实用性的原则上还应考虑成本投入和诊疗效果产出的比值,尤其是面临多种方案可以选择时,应该比较各种方案单位成本投入的效果产出比值或单位效果产出的成本投入比值,或考察增量成本 - 效果比值,选择和推广价廉物美的证据。

五、时效性

临床研究的动力源泉源自临床实践中不断发现的临床问题。不论是对病因的探索、对疾病诊断准确性的渴求,还是对疾病良好结局的期许都会促进临床研究者们不断地围绕临床需要解决的难点和热点问题展开深入研究,以进一步完善和拓展现有的证据,甚至否定现有证据。临床医生正是在不断的发现问题、提出问题和解决问题的过程中,对问题的认识更趋于真实,促进了临床医学的不断发展以及临床医生自身水平的不断提高。比如绝经期妇女激素替代治疗,激素替代治疗(hormone replacement treatment, HRT)在降低潮热、出汗、失眠、情绪改变、生殖道萎缩、尿频尿急和骨质疏松方面的效果已被大量的随机对照试验证实并被临床广泛接受,其在预防心血管疾病、老年痴呆方面的作用,也逐渐被发现。激素替代治疗能降低总胆固醇、LDL,

笔记

提高 HDL 水平等，似血管保护作用，但在一级预防中心脑血管病风险反而增加了 29%，二级预防中非致命性心肌梗死和心脑血管死亡在 HRT 组和对照组之间并无统计学差异，这些后续研究结果让盲目热衷于 HRT 预防心脑血管疾病的热情降温。但证据的应用者们通过分析和总结，发现了研究存在的问题，比如入选的研究对象年龄跨度大，入选年龄 50~76 岁（平均 63 岁），其中 60~69 岁占到 45.3%，从而提出激素应用的开始时间与血管内皮的完整性是否与心脑血管疾病发生的风险相关，进而提出治疗的时间窗问题。另一方面，激素治疗的不良反应也一直困扰临床医生，在绝经期症状改善、骨量丢失减少的同时，也引发了对激素与子宫内膜癌、乳腺癌及卵巢癌发病率上升的担忧，促使人们开始研究激素剂量改变、雌孕激素联合应用等新问题。其实在这一领域中，有关年龄、BMI、家族易感、激素使用的时间、有无心脑血管疾病史都可能影响最后结果，因此来自群体的研究结果在具体病人的应用中会遇到各式各样的问题，一方面可促使临床医生进一步查找更适合的证据，另一方面也促进了相关研究的开展。

第二节　生物学依据

在临床研究中，任何科学研究的设计和实施都应以生物学依据（biologic evidence）为研究基础。在证据的应用中也要考虑临床诊治的生物学依据。例如某一药物治病的效应，要从生物学机制上研究药物的作用环节，为何无效或为何产生不良反应都应有相应的生物学机制。

针对发病机制，有关病原生物学的研究在感染性疾病中具有很重要的意义。比如结核病的病原体是结核杆菌，应用抗结核杆菌的药物能够有效地控制结核病的发展并治愈结核病。病毒性肝炎的病原体是肝炎病毒，乙型肝炎病毒所致的慢性迁延性肝炎与肝癌的发生有很强的关联性，而目前针对乙型肝炎病毒的治疗药物还不尽如人意。肿瘤的发病机制比较复杂，在长期的观察和研究中发现，人乳头瘤病毒（HPV）部分类型的感染与生殖道肿瘤有明显的相关性，比如 HPV16、HPV18、HPV33、HPV52 等感染与宫颈癌的发生有密切关系。针对病毒感染相关的疾病，由于缺乏直接杀灭病毒的药物，主要还是通过研究相关疫苗来预防感染，或通过干扰素诱导体内细胞产生抗病毒蛋白以及调节体内免疫系统来减少和控制病毒。这些都涉及病原体致病与抗病原体药物作用的生物学依据。

从基于人群的流行病学研究或基于临床病人的观察总结，常常会发现不同人种或个体之间存在生物学上的差异。例如黑人的高血压发病率较其他人种高且危害重，临床证明有效的 β-受体阻滞剂、血管紧张素酶抑制剂，对黑人高血压患者的治疗效果，较利尿剂差；同时血管紧张素酶抑制剂引起的血管性水肿显著，较其他人种的发病率高。在丙型肝炎的治疗中也发现，黑人对聚乙二醇干扰素 - 加利巴韦林治疗的持续病毒应答（sustained virus response，SVR）率较白种人低。家族遗传性也体现在疾病的发病率和对药物的反应性上的差异。比如乳腺癌、卵巢癌和结肠癌有家族聚集的倾向，发病的风险高于普通人群。因此在疾病防治中要注意人种、民族和遗传因素的生物学特点。

临床治疗中还观察到，使用抗生素治疗感染性疾病过程中，同一种致病菌存在对同一种抗生素的敏感菌株和耐药菌株，或在抗生素不断使用中渐渐产生对抗生素耐药的现象。在肿瘤病人的化疗治疗中，同一种组织来源的肿瘤，也观察到化疗敏感性或耐药性的差异。这些差异的存在和产生都需要研究者去探索其相应的生物学机制，从细胞甚至分子水平上予以解决，以期为患者制订个体化的诊疗方案。比如，对于丙型肝炎的治疗，经过长期的研究，将传统的干扰素聚乙二醇化，使其被包裹后阻止了蛋白酶的水解，减少了分子的免疫原性，大分子干扰素还防止了肾小球的滤过，使之比普通干扰素更快地达到血药高峰浓度，改变了普通干扰素不稳定的血药浓度高峰低谷现象，并能维持较长的时间。聚乙二醇化的结果提高了丙型肝炎治疗的 SVR 率，这是丙型肝炎治疗上的一大进步。白介素 -28B（IL-28B）基因编码免疫应答调节器，近

笔记

年来研究发现 IL-28B 存在基因多态性，并与丙型肝炎治疗的效果密切相关：rs12979860 纯合子基因型（CC）患者对标准治疗法的 SVR 率是基因携带阴性（TT）患者的 2 倍。在不同人种 CC 概率差异很大，东亚人最高，非洲裔美国人最低。蛋白酶抑制剂 boceprevir 和 telaprevir 已被通过用于治疗丙型肝炎，在 1 型丙型肝炎病毒感染的 CC 患者，与传统的聚乙二醇干扰素 - 加利巴韦林（PR）二联用药的标准方案比较，加用蛋白酶抑制剂的三联药物治疗方案的 SVR 率为 90% 和 70%，而 TT 患者为 70% 和 30%，提示 IL-28B 指导下选用三联或二联药物治疗方案，可能获得最佳成本 - 效果。rs8099917 纯合子 GG 则是丙型肝炎治疗的不良预后基因。等位基因 CC 和 GG 的发现，为丙肝治疗个体化方案的制订提供了可遵循的生物学依据。

随着人类基因计划研究的深入，对于个体基因遗传特征与疾病的临床表象的关系探讨，尤其是单核苷酸多态性与疾病发生及临床用药敏感性的预测，可能为今后临床研究和诊治提供更多的个体方面的生物学依据。

第三节　病理生理学依据

临床实践中，疾病的发生、病情轻重、并发症发生、结局以及临床干预措施对疾病的影响都有相应的病理生理变化和特点。临床研究中若纳入不同病理生理状态的研究对象，不仅临床表现千差万别，对诊疗措施的反应也大相径庭。在证据的应用中要考虑患者的病理生理状态以及治疗过程中的病理生理反应。例如，曾在我国流行的 SARS，部分中毒症状及肺部炎变损害明显、呼吸功能明显障碍者，适时适量的应用肾上腺皮质激素治疗，取得了较好的临床效果，挽救了不少病人的生命。但对于中毒症状以及肺部损害不严重、呼吸功能影响较轻的患者，应审慎应用激素，避免因给药带来不良后果（如股骨头坏死）。

任何最佳证据的应用，一定要仔细考虑患者由于病理生理损害不同而导致临床病情的复杂性。临床研究的成果，由于样本量的限制，就研究对象而言，相对单纯同质，因而即使是最佳证据，对临床的总体代表性而言，往往有其局限性。如果患者病情与研究对象存在差异，比如病情轻重不一致，年龄、性别或并发症有差异，就有可能在机体功能、代谢及药物的反应方面都存在不同，因而证据的应用就不宜简单照搬。例如：对轻、中型高血压患者降压的随机对照试验研究，证明了钙离子通道阻滞剂及血管紧张素酶Ⅱ受体阻滞剂有良好的降压效果，如果面对的患者是一位 3 级高血压患者（BP≥180/110mmHg），或者伴有心脑器官损伤或有其他并发症或合并症等的高危患者，单一降压药物应用就不能达到有效降低血压并预防心脑血管并发症的预期效果。因此，应注意被引用的证据中，纳入的研究对象有无 1、2、3 级高血压或低、中、高危险分层，如有 3 级或伴有相关并发症的高危患者，且与该患者病情相似，那么相应的证据就有被采纳的价值。否则，另当别论。

在估价疾病预后和对有关危险因素的干预，借以改善患者预后、提高患者的生存质量，这就要求将多种相关因素的研究证据，结合患者疾病的病理生理特点进行具体分析和评价，以估计可能发生某一事件的概率并提出有针对性的干预方案，从而防止或降低不良事件的发生，以达到改善患者预后的目的。

第四节　社会 - 心理及经济特点

在国家层面上，鉴于不同国家和地区，社会制度、人文和自然环境、经济发展水平和人群的受教育程度存在着极大的差异，这些差异不仅体现在临床研究的切入点和实施不同，也会在证据的推广应用中存在差异。例如：10 年前，发达国家在新药和新的干预措施的评价中，增量成本≤50 000 美元 /QALY，即作为可以接受的标准。而在发展中国家和贫困国家，每增加

1QALY，将多付出 40 万元人民币的成本，无论从国家层面、社会层面或个人层面，负担都难以承受。10 年以后，美国的标准已达到 100 000 美元 /QALY。按照 2011 年世界银行的统计，美国人均 GDP 达到 48 442 美元，而中国人均 GDP 为 8466 美元，100 000 美元约为美国人均 GDP 的 2.1 倍，为中国人均 GDP 的 12 倍。根据 WHO 药物经济学的推荐意见（2010）：增量成本 - 效果比（incremental cost-effective ratio，ICER）< 人均 GDP，增加的成本完全值得；人均 GDP < ICER < 3 倍人均 GDP，增加的成本可以接受；ICER > 3 倍人均 GDP，增加的成本不值得。所以引用发达国家的标准，应慎重。再者不同发展水平的国家和地区，所面临的疾病负担和急需解决的问题也有很大不同，例如宫颈癌是发展中国家妇女常见的恶性肿瘤之一，极大地危害着女性健康。宫颈脱落细胞学检查已经成为经典的宫颈癌筛查手段，并被全世界广泛接受和采纳。但在一些贫困国家和地区，由于资源匮乏和技术人员缺乏，醋酸试验或碘着色通过肉眼观察被用来取代细胞学筛查。而且宫颈癌筛查的宣传和人们对其重要性的认识以及传统观念影响使该方法的接受程度与发达国家相比，也大大落后。在这样的地方，如果再推行 HPV 检测联合宫颈脱落细胞学筛查宫颈癌及癌前病变，就显得勉为其难了。相反，发展中国家根据当地的经济发展水平、宗教和文化现状，针对一些对当地人群健康危害大的疾病如急性传染病、呼吸道感染、腹泻、营养不良、地方病等的防治性研究，或者诸如上述的宫颈癌肉眼筛查法等行之有效的证据，在发达国家可能就没有如此重要的价值。

医生的知识水平是循证医学实践不可或缺的要素。能充分运用临床基本知识和积累的经验，结合具体病人的实际问题，搜寻和查找新近最佳证据为患者诊治服务，这是一个优秀临床医生的职责所在。医生是否具备循证医学实践的知识，能否在众多的研究结果中筛选到最佳证据，是否具有对先进研究成果的敏锐观察力，都关系着循证医学实践的最终效果。

在最佳证据的应用中，患者的意愿（preference）也不应被忽视。病人的依从性受自身文化水平、社会经济状况、价值观甚至宗教信仰的影响。了解患者的心理及所处的社会经济状态，倾听患者对治疗的意见，尊重他们的价值观，给予患者相关的医学知识教育，加强医患交流，消除患者可能存在的顾虑，尽可能地给患者提供方便的服务，并与患者建立良好医患关系，提高他们的依从性，让患者愿意接受最佳证据指导的治疗，是实践循证医学的基础。

第五节　证据的应用要权衡利弊

最佳证据是否可以用于指导个体患者的医疗决策，一个显而易见的问题是：拟采用的诊治措施能给患者带来多大的利益，同时还要考虑它们被应用后可能产生哪些不良反应及其对患者造成的危害程度。所谓“利”是指临床干预措施给患者带来的临床治疗获益，即疾病缓解的程度、治愈的可能性等，而“弊”则指不良反应的大小及危害程度。临床经济学的角度还要考虑成本 - 效果比值。所以在证据的应用中要客观评估临床诊治措施对患者的利与弊，选择利大于弊的证据。

临床意义或价值有很多评价和表述的指标，一些量化的指标更有利于医生间或医患间的交流。下面为不同研究领域的常用指标：

病因及危险因素研究证据：常采用事件发生率（event rate，ER），相对危险度（relative risk，RR），归因危险度（attribute risk，AR），病因学分数（etiologic fraction，EF），优势比（odds ratio，OR），绝对危险增高度（absolute risk increase，ARI），相对危险增高度（relative risk increase，RRI），暴露多少研究对象可导致一例发病（number needed to harm，NNH）。

诊断学研究证据：采用的指标有敏感度（sensitivity，SEN），特异度（specificity，SPE），准确度（accuracy），预测值（predictive value，PV），似然比（likelihood ratio，LR）。

治疗及干预性研究证据：常用指标包括治愈率，有效率，病死率，绝对危险降低度（absolute

笔记

risk reduction，ARR），相对危险降低度（relative risk reduction，RRR），治疗多少患者可获得一例效果（number needed to treat，NNT）。

临床经济学研究证据：根据研究的目的、待评价的效果可选择成本最小化分析（cost-minimization analysis，CMA）、成本 - 效果分析（cost-effectiveness analysis，CEA）、成本 - 效益分析（cost-benefit analysis，CBA）和成本 - 效用分析（cost-utility analysis，CUA）。其量化指标有最小成本、成本 - 效果比、成本 - 效益比和成本 - 效用比以及增量成本 - 效果比等。

临床研究发生的不良事件，尤其是与所采取的诊疗措施相关的不良反应，如药物不良反应（adverse drug reaction，ADR），要作为证据的“弊”加以关注。在循证医学个体化实践中，药物不良反应的严重程度（比如是否对脏器功能造成损伤及损伤的程度，损伤是否可逆），重要事件（致残、致死）率的高低都要认真思考。其他指标如：治疗多少病例才发生一例重要的不良事件（NNH），预防或治疗药物不良反应或重要不良事件所投入的成本是多少，以及由此带来的增量成本 - 效果比（incremental cost-effective ratio）等。

对研究证据利弊评价的最终目的是要为自己面对的具体患者的诊治做出决策，根据利弊比（likelihood of being helped vs harmed，LHH）来衡量证据应用可能为患者带来的最佳效果和不良反应。比如某种防治措施的 NNT 为 10，NNH 为 30，即此种措施的利弊比（LHH）= 1/10∶1/30，约为 3∶1。意为该措施对患者带来的受益 3 倍于危害。但这一结果（利大于弊）并不是决策的最终依据，还要进一步结合专业及临床实际，考虑患者对结果的意愿及价值取向来校正 LHH，并最后决定是否采用该措施。表 13-2 可通过获得的 NNT 及 NNH 查得利弊比。

表 13-2 干预效果的利弊比（LHH）

比值		1/NNT									
1/NNH		1/10	1/20	1/30	1/40	1/50	1/60	1/70	1/80	1/90	1/100
	1/10	1.0	0.5	0.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1
	1/20	2.0	1.0	0.66	0.5	0.4	0.34	0.3	0.3	0.2	0.2
	1/30	3.0	1.5	1.0	0.75	0.6	0.5	0.4	0.4	0.3	0.3
	1/40	4.0	2.0	1.3	1.0	0.8	0.7	0.6	0.5	0.4	0.4
	1/50	5.0	2.5	1.7	1.3	1.0	0.9	0.7	0.7	0.6	0.5
	1/60	6.0	3.0	2.0	1.5	1.2	1.0	0.8	0.8	0.8	0.6
	1/70	7.0	3.5	2.3	1.8	1.4	1.2	1.0	0.9	0.8	0.7
	1/80	8.0	4.0	2.6	2.0	1.6	1.4	1.1	1.0	0.9	0.8
	1/90	9.0	4.5	3.0	2.3	1.8	1.5	1.3	1.2	1.0	0.9
	1/100	10	5.0	3.3	2.5	2.0	1.7	1.4	1.3	1.1	1.0

第六节 个体化干预的效果预测

“最佳证据”应用于循证医学实践对个体患者之干预，究竟会产生什么效果？可能达到什么结局？这是患者和医生都十分关注并期望回答的实际问题。

患者，男性，70 岁，患高血压病有 20 年病程，一直在用药物进行治疗，但血压维持在 150/90mmHg 左右，疗效欠满意。3 年前发现糖尿病（2 型），接受胰岛素治疗，血糖维持在 6.0~6.5mmol/L，且血脂代谢紊乱，甘油三酯 2.5mmol/L，血胆固醇 6.5mmol/L，HDL-Ch 0.8mmol/L，LDL-Ch 3.8mmol/L。肝、肾功能正常。

笔记

患者有18年的吸烟史，现10支/日。戒烟困难。

查体：体型稍胖，身高172cm，体重74kg，BP 150/94mmHg，心脏无异常体征，双肺正常，肝脾不大，下肢不肿。心电图正常。

患者就医的目的：①询问根据现况发生冠心病或脑卒中的危险度有多大？②如何进行有效的治疗以降低发生心脑血管病的危险性？

根据数十年来的大量的流行病学研究结果所获得的可靠证据，明确了心脑血管疾病发病的主要危险因素，见表13-3。

表13-3　心脑血管疾病的危险因素

高血压（1~3级）	早发心血管疾病家族史（一级亲属年龄发病<50岁）
年龄（男性>55，女性>65）	腹型肥胖（腰围：男性≥90cm，女性≥85cm）或肥胖（BMI≥28kg/m²）
吸烟	
糖耐量受损和（或）空腹血糖异常	
血脂异常	

（摘自《中国高血压防治指南》第3版）

从表13-3所列的心脑血管疾病的危险因素中，本病例具有发生心脑血管病的危险因素为：高血压、吸烟、血脂紊乱、糖尿病、年龄>55岁共5项。虽然患者现在尚无心脑血管病的临床证据，但发生的概率在未来是颇高的。然而这仅仅是定性的结论。是否有定量化的预测指标呢？如果有，那么进行有针对性的干预后是否可以表明发生心脑血管病的危险度会降低呢？

现在让我们分别引用中国医学科学院阜外医院国家"十五"攻关课题总结的有关缺血性心血管疾病（ICVD）综合危险度评估工具图13-1，可为本例患者进行评估。

根据图13-1，该患者年龄>59岁计4分，SBP计2分，体质指数（BMI：$74kg/1.72m^2=25$）计1分，胆固醇计1分，吸烟计2分，糖尿病计1分，总分为11分。从第三步图中可查出未来10年发生ICVD的危险度为12.8%。

国外的定量评估ICVD工具中，也颇有重要参考价值，同样可用于对尚未发生缺血性心血管疾病的患者的预测，或作干预后估价危险度下降的效果预测，这里，要推荐新西兰高脂血症及高血压管理指南中所列"估计心血管疾病的危险与治疗益处（estimating cardiovascular risk and treatment benefit）"。在该资料中，确定心血管事件为：新发的心绞痛、心肌梗死、冠心病死亡、卒中或短暂性脑缺血发作、充血性心衰或外周血管综合征。

研究者根据主要危险因素的研究结果来预测未来5年内发生上述心脑血管事件的危险度。如经过干预后，视这些可控危险因素的变化，以确定治疗将带来的益处。

为了形象化的指导干预实践，他们绘制了多维图（见彩图13-2）ICVD预测危险度与干预效果图。

彩图13-2中主要变量为：性别、年龄、有无糖尿病、吸烟与否、血压水平、总胆固醇/高密度脂蛋白胆固醇（HDL-Ch）的比例，共6大项。图中以不同颜色的小方块图示所查病例发生ICVD的危险程度。

在判断危险度及其治疗获益方面，列出表13-4。

基于总胆固醇下降20%或SBP下降10~15mmHg或DBP下降5~10mmHg，估计CVD事件5年约下降1/3。

笔记

缺血性心血管疾病(ICVD)综合危险度评估工具

男性

第一步：评分

年龄(岁)	分值
35～39	0
40～44	1
45～49	2
50～54	3
55～59	4

收缩压	分值
<120	－2
120～129	0
130～139	1
140～159	2
160～179	5
≥180	8

体质指数	分值
<24	0
24～27.9	1
≥28	2

总胆固醇	分值
<200	0
≥200	1

吸烟	分值
否	0
是	2

糖尿病	分值
否	0
是	1

第二步：求和

危险因素	得分
年龄	______
收缩压	______
体质指数	______
总胆固醇	______
吸烟	______
糖尿病	______
总计	______

10年ICVD绝对危险参考标准

年龄	平均风险(%)	最低风险(%)
35～39	1.0	0.3
40～44	1.4	0.4
45～49	1.9	0.5
50～54	2.6	0.7
55～59	3.6	1.0

第三步：绝对危险

总分	10年ICVD危险(%)
≤－1	0.3
0	0.5
1	0.6
2	0.8
3	1.1
4	1.5
5	2.1
6	2.9
7	3.9
8	5.4
9	7.3
10	9.7
11	12.8
12	16.8
13	21.7
14	27.7
15	35.3
16	44.3
≥17	≥52.6

女性

第一步：评分

年龄(岁)	分值
35～39	0
40～44	1
45～49	2
50～54	3
55～59	4

收缩压	分值
<120	－2
120～129	0
130～139	1
140～159	2
160～179	3
≥180	4

体质指数	分值
<24	0
24～27.9	1
≥28	2

总胆固醇	分值
<200	0
≥200	1

吸烟	分值
否	0
是	1

糖尿病	分值
否	0
是	2

第二步：求和

危险因素	得分
年龄	______
收缩压	______
体质指数	______
总胆固醇	______
吸烟	______
糖尿病	______
总计	______

10年ICVD绝对危险参考标准

年龄	平均风险(%)	最低风险(%)
35～39	0.3	0.1
40～44	0.4	0.1
45～49	0.6	0.2
50～54	0.9	0.3
55～59	1.4	0.5

第三步：绝对危险

总分	10年ICVD危险(%)
－2	0.1
－1	0.2
0	0.2
1	0.3
2	0.5
3	0.8
4	1.2
5	1.8
6	2.8
7	4.4
8	6.8
9	10.3
10	15.6
11	23.0
12	32.7
≥13	≥43.1

单位：收缩压 mmHg　体质指数：kg/m^2　总胆固醇：mg/dl　空腹血糖 mg/dl

图 13-1 缺血性心血管疾病(ICVD)综合危险度评估工具

应用这份图表，查看该病例5年内发生CVD的几率。该例为男性，有糖尿病、吸烟、血压150/94mmHg、TC/HDL＝8.1。查彩图13-2显示为红色小块，表示5年内发生CVD者＞30%。高危性，应予积极治疗，如果能有效降低血压，控制糖尿病及高脂血症和戒烟，则可望发生CVD危险性会大大下降。其降低水平仍可查图确定，这将取决于对病人的干预及其效果。

临床医学的研究工作，就是要获得对防病治病有价值的成果。然后又将其应用于防病治病的具体实践，在实践中进一步评估临床效果，并从具体的患者个体应用中发现新问题，以此不断提高临床诊治的水平，达到保障人民健康的最终目的。循证医学临床实践应遵循个体化的原则，以最终解决患者个体的重要实际问题。

表 13-4　心血管疾病(CVD)的危险度和治疗获益方面估价表

危险水平（5年发生致死与非致死CVD）		益处(1)（5年治疗100人可防御CVD）	益处(2)（治疗5年的NNT）
非常高	红色 >30%	>10%	<10
	深橙色 25%~30%	9%	11
	淡橙色 20%~25%	7.5%	13
高	黄色 15%~20%	6%	16
中	绿色 10%~15%	4%	25
低	蓝色 5%~10%	2.5%	40
	紫色 2.5%~5%	1.25%	80
	淡蓝 <2.5%	<0.8%	120

（方　芳　王家良）

笔記

第十四章　决 策 分 析

决策分析是通过决策模型再现问题、利用概率和结局估计值等，帮助确定最佳行动方案的一个过程。决策分析目前已有一套基本的理论与方法，如 Bayes 决策法、决策树、Markov 模型等；正确决策的前提是充分掌握准确的、有价值的信息和正确的预后预测。决策类型一般分为确定型决策、风险型决策和不确定型决策，其中绘制决策树、各事件的发生概率和结局的定量估计是风险型决策的三大要素；敏感性分析、阈值分析和决策追踪是确保决策正确的重要措施。尽管临床决策分析的方法尚不完善，但决策的思想及其思维方式，对于临床医生来说，仍至关重要。

第一节　概　　述

决策(making decision)是基于不确定性的问题，通过一些方法与手段，从众多备选方案中选定最优方案的一个过程。决策分析(making decision analysis)则是通过决策模型再现问题、利用概率和结局估计值等，帮助确定最佳行动方案的一个过程。而临床决策分析(clinical decision analysis，CDA)是指由医务人员参与实施的、针对疾病的诊断和防治过程中风险与获益的不确定性，通过查阅文献资料，充分掌握证据，特别是在掌握最新最佳证据的基础上，结合以往临床经验和患者的实际情况，分析比较两个或两个以上可能的备选方案，从中选择最优者来进行临床实践的决策过程。

在临床实践中，当掌握了患者的全部情况、明确了需要解决的临床问题，以及找出了解决这些问题的最佳证据之后，那么就要对患者的诊治处理进行具体的临床决策分析，而后做出正确的决断。决策正确与否取决于 3 个前提条件：备选方案是否齐全、各事件的概率估计是否准确、结局的定量是否合情合理。无论是对概率还是结局的定量估计，都应来自证据充分的临床研究或文献资料。因此，决策前，应开展相应的循证医学研究，通过广泛调查、文献复习、科学分析，以充分掌握基础信息，进而选用论证强度高的结果评价备择方案，或应用准确的概率来描述事件的不确定性，以及使用合理的效用值来定量描述结局，然后计算各备择方案的期望值，从中选择收益最大者，应用于临床实践。

进行临床决策的难点在于缺乏各种决策证据，如干预后出现的各种结局事件常不确定，临床事件的发生不以人的意志为转移，虽可预测，但发生概率难以确定；同时鉴于各种备选方案都有优缺点，需要权衡利弊后加以取舍，也往往令决策者左右为难。

一、临床决策的模式

(一) 患者做主模式

也称单纯知照模式(pure informed model)，是指由患者自己做出选择。即医生只是提供各种方案的利弊等相关信息，患者根据自身的经验以及理解独立做出选择。在这里，患者是唯一的决策者，医生只提供客观信息而不参与决策。

(二) 医生做主模式

也称家长作风模式(paternalistic model)，是由医生为主导做出决策。即医生单独或者和其他医生一同考虑方案的利弊后而做出选择。在这种情况下，患者不参与决策，但前提是：假定医生知道哪种方案对患者最为适合。

笔记

（三）共同决策模式（shared decision model）

共同决策模式是指在决策过程中医生与患者保持互动、共同做出决策。这种模式的前提是信息充分对称，即医生向患者提供病情相关的信息，包括各种检查和可能的治疗方案以及各方案的利弊等情况，患者则如实提供自身的病情、生活方式以及意愿和价值取向等信息。然后医患双方再对相关的诊治方案进行讨论，并结合实际情况（如社会、家庭、医院技术条件等因素）以寻求一个最优的诊治方案。这也是循证医学所倡导的最佳临床实践模式。在这种模式下，医患关系融洽，患者的意见得到充分的尊重。医患双方都是决策者，缺少任何一方，都是不完整决策。

二、临床决策的具体分型

（一）确定型决策

确定型决策是指事件的结局已经完全确定的情况下，只需通过分析各种方案的最后得失，做出选择。决策方法有增量分析法，即通过计算方案之间的增量收益与增量成本进行决策的方法。

（二）风险型决策

风险型决策是指事件发生何种结局不能肯定，但其概率可以估计，决策带有一定的风险性。风险型决策有 3 个前提条件：即两种以上的结局、自然状态下事件的概率能够估计、不同结局的利弊可以估算。方法包括期望值决策法、Bayes 决策法。

（三）不确定型决策

是指事件的结局及其事件发生概率均不确定的情况下进行地决策。决策时可参考一些准则，如乐观准则、悲观准则、后悔值准则等，进行定性判断。

三、临床决策过程

（一）明确需决策的问题

首先要考虑，面临的是什么决策问题？要达到什么目标？有哪些备择方案？存在哪些影响因素？需要收集哪方面的信息，等等，总之，对涉及该问题的方方面面要全面把握，为最终正确解决问题打下良好的基础。

（二）决策问题的组织和构建

在选题的基础上，分析问题的组成要素，理清要素之间的关联，预测事态发展的趋势，通过一些直观的方法把各种方案实施的进程、路径、结局等形象地表现出来，如常用的决策树法。

（三）系统全面搜集信息

决策的前提是掌握足够的决策信息，这是决策中最为核心的一步。应系统全面搜集包括各种可能事件的概率、结局的量化指标（生存率、生存质量、成本等）等在内的决策信息。

（四）决策分析是解决问题的阶段

利用一些公式或模型进行分析和运算，得到各种方案的期望值供决策者参考。

（五）敏感性分析

鉴于各种事件的发生概率或结局效用值，有时难以确定，可以选用不同的概率或效用值重新进行分析和运算，分析当前决策结果的稳健性。

第二节 概率与效用值的估计

决策分析属定量分析，需要对临床问题进行量化处理，即用概率来定量临床事件的不确定性，用生存率、生存质量、成本或效用值等来定量结局。

笔记

一、概率的估计

概率是对事件结局发生的机会大小或可能性的定量估计。就临床问题而言，概率可通过以下途径获取：一是有针对性地进行临床研究，如开展随机对照临床试验等；二是通过二次文献研究，如系统评价等方法获得较为理想的概率估计。

其中基线估计(baseline estimate)是对临床事件发生概率或其他参数的最佳估计。例如：急性心肌梗死患者的住院病死率为10%左右，Ⅲ期鼻咽癌患者的5年生存率为64%左右。事件发生的概率既可以是一个点估计值，也可以是一个区间估计，以区间最为常用，区间值可以是多个研究结果的变化范围，也可以是某个研究结果的可信区间，区间范围越宽，表明对事件发生概率估计的把握性越差。由于决策相关的临床数据多来源于文献资料(具体参见本教材第五章)，在此简述如何处理来源于文献资料的数据，对于选中的文献资料：①若数量较多，既可取其结果的平均值，也可确立变化范围，一般以区间值较为合适；或者，若研究结果的同质性较好，可用meta分析，计算加权平均值。②倘若文献很少，有时仅能采用一篇，则可取其结果的可信区间(如95% *CI*)。

从文献获取的临床事件概率值应审慎对待，决策时应考虑：①研究的人群是否具有同质性：文献中的人群特征，如性别、年龄构成以及人种等是否特殊；②干预的临床手段是否相同或相近。这些都涉及文献的再次评价，尽量采用高质量文献的结果，这对正确决策至关重要。另外一个极端情况是，无法从现成文献中找到相关数据，只能借鉴专家意见。

二、效用值的估计

这也是决策分析的难点之一，由于对某种临床结局的效用度量存在较大的主观性，不可避免地出现较大的差异。对不同结局，如何赋予相应的效用值、比较真实地反映结局严重程度和患者的意愿呢？结局效用的常用量化方法有：画线法，博弈法，时间权衡法等三种。下面举例说明：某男性患者，68岁，有糖尿病病史多年，现因意外出现左脚刺伤，出现严重感染并可能出现坏疽。该患者面临两种选择，一是截肢，以保全性命，但失去左脚；二是保守治疗，但是存在感染和坏疽扩散的危险，从而有两种可能结局：痊愈或死亡。

(一) 画线法(visual analog scale)

如定义死亡的效用值为0，健康生存(痊愈)为1，介于0与1之间的则代表各种不同状态的效用值，画一条直线如长10cm，如下，让患者在认为可以接受的地方标记，表示可以接受某结局所对应的效用值。假如失去左脚时，该糖尿病患者在直线上画出相应的效应值(相当于正常生活能力的质量)为0.8，倘若能以此状态生存10年，那么他的质量调整寿命年为$10 \times 0.8 = 8$ *QALYs*。

0 ———————————————— 1.0

(二) 博弈法(standard gamble)

效用值还可以借用决策树表示(图14-1)，若告诉患者，保守治疗的痊愈可能性$P = 0.9$时，患者可能会毫不犹豫的选择保守治疗，此时若改变P值，当$P = 0.6$时，患者可能要选择截肢。

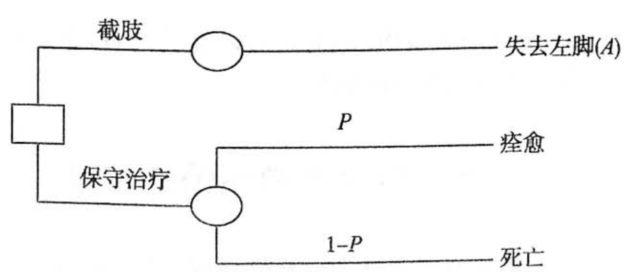


图14-1 截肢的决策树

不断改变 P 值询问患者，当达到某一数值如 $P=0.8$ 时，患者认为两种方案差不多，此时的 P 值称为均衡概率(breakeven probability)，其数值就是结局 A 的效用值，即 $U_A=0.8$，意思是，如把痊愈视为 1，死亡为 0，患者把失去左脚认为相当于 0.8。如果最差的状态不是死亡，而是一种较 A 更差的结局 B，见图 14-2。

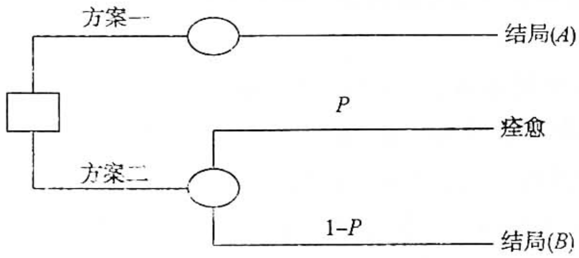


图 14-2　保守治疗决策树

同样用上述方法可以确定 A 的效用值 $U_A=P+(1-P)\times B$。

（三）时间权衡法(time trade-off)

某种结局(A)被认为好过死亡，患者面临两种选择：①结局 A 生存 t 年后死亡；②健康状态生存 $X(<t)$ 年后死亡。如图 14-3 中，横轴表示时间，纵轴表示健康状态；不断变化 X，直到患者认为两种选择无差异时，此时结局 A 的效用值 $U_A=X/t$。

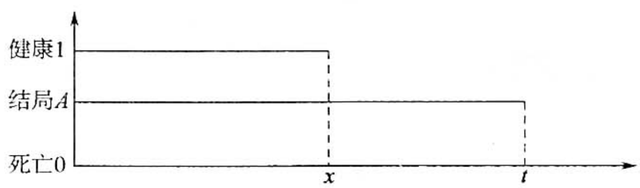


图 14-3　终点为死亡时的权衡

如果患者最后不是死亡，而是两种结局维持一段时间后恢复健康，那么，两种选择是：①暂时结局 A 延续 t 年后恢复健康；②暂时结局 B 延续 X 年后恢复健康。如图 14-4 中，同样不断变化 X，直到患者认为两种选择无差异时，此时结局 A 的效用值 $U_A=1-(1-B)X/t$。

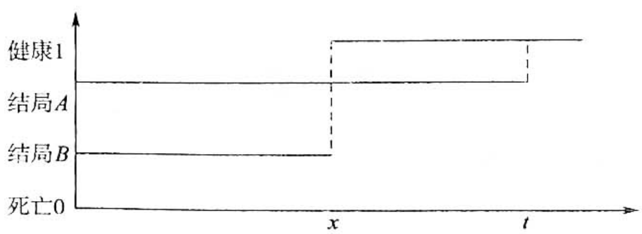


图 14-4　终点为健康时的权衡

第三节　决策树模型

决策树模型是利用决策树来描述各种决策方案在不同自然状态下的收益，据此计算各方案的期望收益而做出决策。决策树(decision tree)按逻辑、时序把决策问题中的备择方案以及相应结局有机组织起来并用图形展示，如同一棵从左至右不断分枝的树，包括一些结点与分枝：决策结点(decision node)，用小方框表示，由此结点发出的方案要求决策者从中做出选择，由决策结点发出的分枝叫决策枝；机遇结点(chance node)，用小圆圈表示，由此结点发出的事件不受人

笔记

的意志所控制，是随机的，但其概率可以估计，由它引出的分枝叫机遇枝或概率枝。机遇节点的期望值为此节点下各分枝概率(P)与结局值乘积之和。

决策树的构建原则是要求"简单且全面"，简单是要求所建立的决策树合乎逻辑，清楚明了，这是从决策树的形式方面来说的；全面则是要求决策树所反映的问题充分考虑到各种情况，这是从内容上考虑的。决策者必须要考虑到每一种选择的结局或者是后果有哪些可能情况，这是形成一个良好决策的基础。

例 14-1 一名 60 岁男性患者，职业是卡车司机，吸烟、肥胖、有心脏病风险因素（包括高血压和高胆固醇）。因患有严重颈部疼痛，他希望医生给他开 COX-2 NSAID（非甾体抗炎药）治疗颈部疼痛（他听说这种药物适用于颈肩部疼痛），但被拒绝，这是因为最新的证据表明 COX-2 NSAID 能够引发严重心脏不良事件。医生建议采用颈部按摩疗法，而患者又不接受，因为他最近读到一个关于颈部按摩后脑卒中瘫痪的报道，脑卒中瘫痪的风险是他最担心的，此外，药物治疗更方便。医生指出，颈部按摩导致脑卒中的可能性非常小，心脏病发作的危害同样严重。医患双方均陷入两难的困境：到底是颈部按摩治疗（治疗中可能发生脑卒中瘫痪），还是 COX-2 NSAID 治疗（服药中可能发生心脏病）。此时可以通过绘制决策树加以解决。先绘制颈部按摩决策树（图 14-5）。

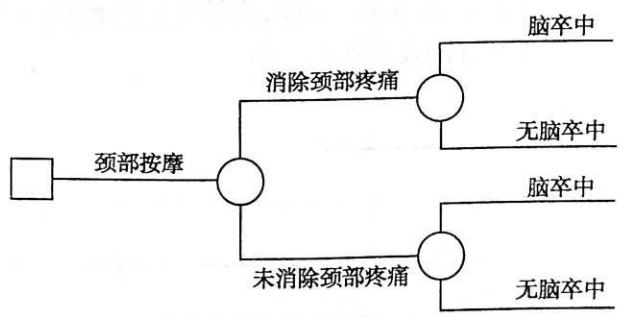


图 14-5 颈部按摩决策树

如果选择 COX-2 NSAID 治疗，该药物虽然有益于解决问题，消除颈部疼痛，但是却有服药期间发生心肌梗死的风险，选择 COX-2 NSAID 的决策树见图 14-6。

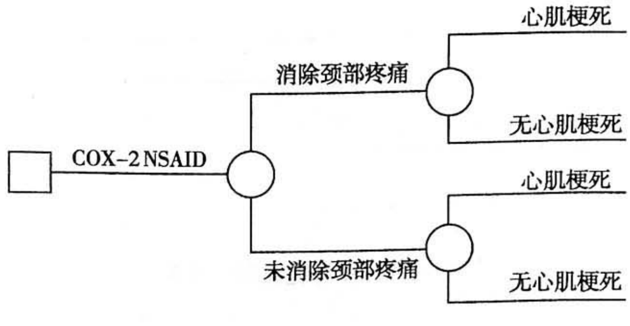


图 14-6 COX-2 NSAID 治疗决策树

合并上述两种选择的决策树见图 14-7。

至此，决策树已清楚展示所要决策的问题，下一步就是收集信息并进行分析。

本例决策所需的基线估计值，假设如下：颈部按摩可有 80% 的机会消除颈部疼痛，其中有 $1/10^6$ 的患者可能会发生脑卒中，其效用值为 0.45，未发生脑卒中患者其效应值为 1.00；而选择颈部按摩但未消除颈部疼痛的患者，有 $1/10^6$ 可能会发生脑卒中，其效用值为 0.30，未发生脑卒中的效用值为 0.80。将这些估计值在决策树中标出（图 14-8）。

笔记

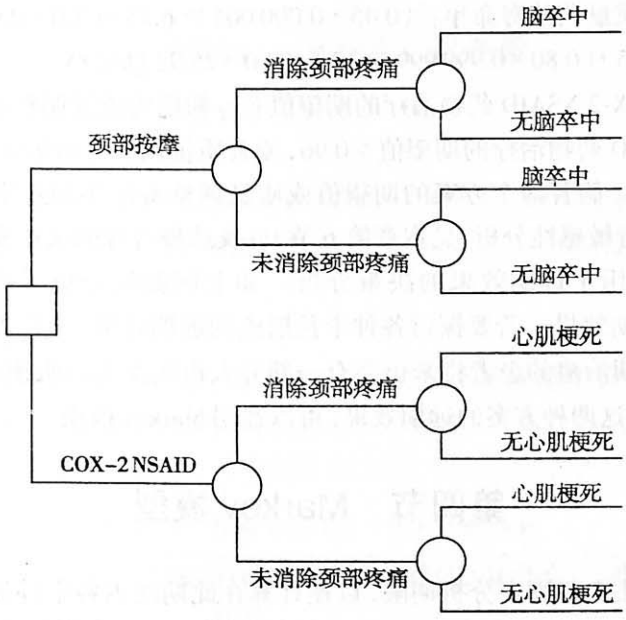


图 14-7　颈部按摩与 COX-2 NSAID 治疗的决策树

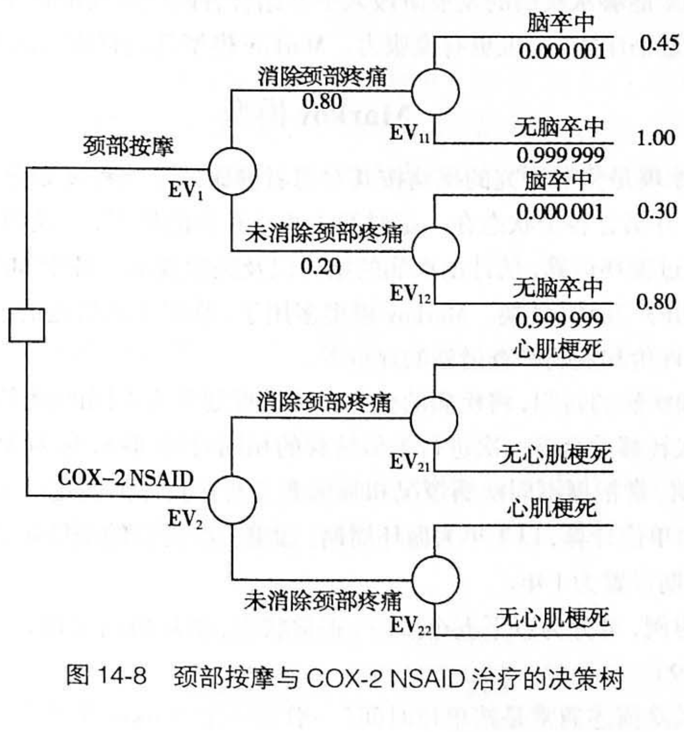


图 14-8　颈部按摩与 COX-2 NSAID 治疗的决策树

本例中还应用了质量调整寿命年(*QALY*)来衡量结果，设定患者无论是“颈部疼痛状态”下生存(无脑卒中)还是有病生存(伴脑卒中状态)，期望寿命都是 15 年。上述四种状态的 *QALY* 分别为 6.75 年、15 年、4.5 年和 12 年。

在此基础上可进一步计算各机遇节点和各种方案的期望值(expected value, EV)。各机遇结点的期望值为该结点各分支概率与结局值的乘积之和，若使用期望质量调整寿命年，则为该结点各分支概率、结局值、*QALY* 值的三者乘积之和。

根据上面的决策树可以计算颈部按摩治疗的期望值(EV_1)如下：

$EV_{11}=0.45\times0.000\,001+1.0\times0.999\,999=0.999\,999\,45$

$EV_{12}=0.30\times0.000\,001+0.80\times0.999\,999=0.799\,999\,5$

$EV_1=EV_{11}\times0.80+EV_{12}\times0.20=0.96$

笔记

颈部按摩治疗的质量调整寿命年：$[(0.45\times0.000\,001)\times6.75+(1.0\times0.999\,999)\times15]\times0.80+[(0.30\times0.000\,001)\times4.5+(0.80\times0.999\,999)\times12]\times0.20=25.92\ QALYS$

同样可以计算 COX-2 NSAID 药物治疗的期望值 EV_2 和期望质量调整寿命年。

如果 COX-2 NSAID 药物治疗的期望值 >0.96，或者质量调整寿命年 $>25.92\ QALYS$，说明药物治疗优于颈部按摩。倘若两个方案的期望值或质量调整寿命年相差不大，医生恐难以下定论。解决的方法是进行敏感性分析（见本章第五节）以及选择合理的决策模式（见本章第六节）。

决策树模型一般用于近期效果的决策分析。如上例就仅分析了颈部按摩以及 COX-2 NSAID 治疗期间的短期效果。若要探讨各种干预措施的远期效果，又将如何进行决策分析呢？如本例中即使药物短期治愈的患者将来仍会有一部分人再次发生心肌梗死、死亡，颈部按摩治疗方案同样如此，评价这两种方案的远期效果，可以选用 Markov 模型。

第四节　Markov 模型

传统的决策树方法需要固定分析期限，以便计算在此期间内各事件概率和效用的平均值。事实上疾病发展是一个动态过程，在疾病发展的不同阶段，各事件发生的概率和结局的效用值是随时间的变化而变化的。如慢性疾病的发展通常可分无症状阶段、轻度不适阶段、中度不适阶段、重症阶段。如果能够从疾病的发展阶段入手并结合各阶段转换的概率进行风险和效益的评估，其结果则更接近临床实际，也更有说服力。Markov 模型就是解决方法之一。

一、Markov 模型

Markov 模型的原理是将所研究的疾病按其对患者健康的影响程度划分为几个不同的健康状态（Markov state），并结合各个状态在一定时间内相互转换的概率，以及每种状态下的资源消耗以及健康结果，通过循环运算，估计出疾病的结局以及医疗成本。各个 Markov 状态通常按照临床规律进行彼此间的“互动”转换。Markov 模型多用于：临床干预措施的评价，临床试验结果的外推，药物经济学评价和疾病筛查措施的评价等。

根据研究目的和疾病的转归，将疾病的整个自然过程划分为不同的健康状态，即 Markov 状态。各种状态在一次转移后至下一次进行类似转移的相同时间间隔，称为 Markov 周期（Markov cycle），亦称转移周期，常根据不同疾病情况和临床意义而作具体的设定。如脑卒中、花粉症等发生的频率多以年为单位计算，以 1 年为循环周期；如果文献提供的是每年的发病率、死亡率或致残率等，则循环周期设置为 1 年。

以肾病综合征为例，可分为以下五个阶段：正常状态，微量蛋白尿期，大量蛋白尿期，终末期肾病，死亡（图 14-9）。

对疾病进程的风险描述通常是按单位时间（一般是一个 Markov 周期）进行的。如：处于微量蛋白尿期的患者每年有 5% 进展到大量蛋白尿阶段，循环周期为 1 年。

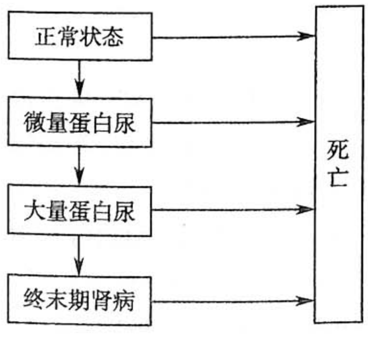


图 14-9　肾病综合征的 Markov 状态转移模型

笔记

二、Markov 模型的分析步骤

与决策树模型类似，Markov 模型也可以分为以下几步：

（一）Markov 状态的设定

将临床问题进行分解，根据研究目的确定各种疾病的状态，从而设定相关的 Markov 状态。对各种疾病状态的逻辑表现形式，可以借助 Markov 树（Markov tree）来表达。各种状态之间的转换必须符合逻辑和临床实际。

（二）基础信息的搜集

即各种状态之间转换概率以及循环周期的确定。主要是从文献中获得，通常用矩阵按照事件发生的逻辑顺序列出来，如上述的肾病综合征的几个状态：正常状态（状态 1），微量蛋白尿阶段（状态 2），大量蛋白尿阶段（状态 3），终末期肾病（状态 4），死亡（状态 5）（表 14-1）。

表 14-1　各种疾病状态 1 年的转移概率

	正常状态	微量蛋白尿阶段	大量蛋白尿阶段	终末期肾病	死亡
正常状态	0.97	0.02			0.01
微量蛋白尿阶段		0.94	0.04		0.02
大量蛋白尿阶段			0.90	0.05	0.05
终末期肾病				0.80	0.20

假设肾病综合征自然状态从第 1 年开始，一个健康人群中（状态 1），每年有 2% 的人进入状态 2，1% 的人进入状态 5；而处于状态 2 的人中每年有 4% 的人进入状态 3，2% 的人进入状态 5；处于状态 3 的人群有 5% 的人进入状态 4，5% 的人进入状态 5；处于状态 4 的人每年有 20% 的人进入状态 5（死亡）。

（三）各种状态效用值的确定

如例 14-2、例 14-3。

（四）Markov 模型的综合分析

如例 14-2、例 14-3。

（五）敏感性分析

同决策树模型一样，Markov 模型也应进行敏感性分析，以判断分析结果的稳健性。

例 14-2　假设某种疾病只有 3 个状态：健康（状态 1）、患病（状态 2）、死亡（状态 3），其转移概率列于表 14-2。

表 14-2　3 种状态 1 年的转移概率

	健康	患病	死亡
健康	0.75	0.20	0.05
患病	0	0.70	0.30
死亡	0	0	1

其意义亦可示意如图 14-10：

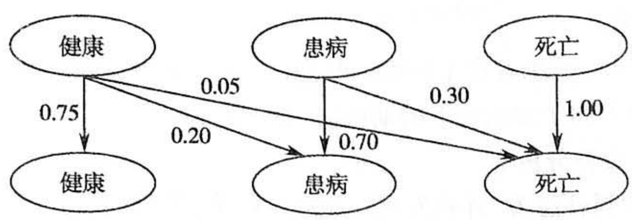


图 14-10　三种状态的转移关系及概率

设定 3 种状态的效用值分别如下：健康为 1，患病状态为 0.5，死亡为 0。构建 Markov 树（图 14-11）。

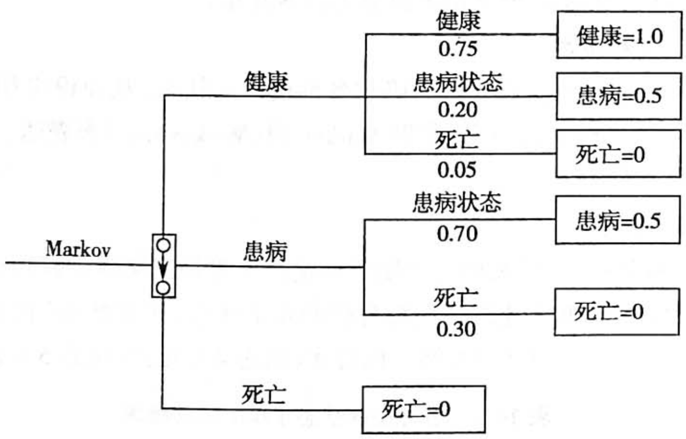


图 14-11 三种状态的 Markov 树

假设有 1000 名符合相关条件的健康人群，从某一时间点开始，依次进入这 3 种状态，据此计算出这一人群此后每年在这 3 个状态中的分布情况，进而计算每一循环周期或者是至研究结束时所有周期累计的资源耗费或各种结局的具体情况。

首先假设 t 为周期（年），a_t 表示第 t 年（周期）时的“健康”人数，b_t 表示第 t 年时的处于患病状态的人数，c_t 表示第 t 年时总的死亡人数。计算如下：$a_0=1000$，$b_0=0$，$c_0=0$；$a_1=0.75\times a_0$，$b_1=0.20\times a_0+0.70\times b_0$，$c_1=c_0+0.30\times b_0+0.05\times a_0$；依此类推：$a_t=0.75\times a_0$，$b_t=0.20\times a_{(t-1)}+0.70\times b_{(t-1)}$，$c_t=c_{(t-1)}+0.30\times b_{(t-1)}+0.05\times a_{(t-1)}$，如表 14-3 所示：

表 14-3 Markov 模型模拟的计算结果

周期(t)	健康(a_t)	患病状态(b_t)	死亡(c_t)	周期(t)	健康(a_t)	患病状态(b_t)	死亡(c_t)
0	1000	0	0	7	133	205	662
1	750	200	50	8	100	170	530
2	560	290	150	9	75	139	786
3	420	315	265	10	56	113	831
4	316	305	379	20	3	9	988
5	237	277	486	25	1	3	996
6	178	241	581	28	0	0	1000

进一步计算每一个循环末的结局，如第 1 个周期末的总质量调整寿命年（$QALY$）为：$750\times 1+200\times 0.5+50\times 0=850$ $QALYs$，依次计算可以获得该模型模拟 28 年的总质量调整寿命年为 5332.7 $QALYs$，而人均质量调整寿命年为 5.33 $QALYs$。这是自然状态下未加干预手段的结果，如果对该人群施加一定的临床干预，则必然影响人群在不同状态下的分布，就可计算在该干预措施下的人均 $QALYs$，如果有多个干预措施，同样可以计算不同措施下的 $QALYs$，从而可以比较不同干预措施的效果。

临床干预措施既可影响患者在不同疾病状态的分布，也可影响各状态间的转换概率，此时可用 Markov 模型评价不同干预措施下的期望值，据此选择适当的干预手段，同时也可结合资源消耗估计，进行成本 - 效益分析等。

例 14-3 某 50 岁男性患者，体检发现左颈动脉硬化，但目前没有任何临床症状。对于该患者的治疗，若不处理则有较高的发生脑卒中的风险，鉴于目前无任何症状，临床上可以有两种

笔记

选择，一是暂时临床观察，二是行颈动脉内膜切除术。若只进行临床观察，虽可避免手术相关的短期危险因素（围手术期死亡，手术中发生脑卒中），维持无症状性颈动脉硬化状态（在一段特定时间内不发生脑卒中），但面临将来发生脑卒中的高风险。假设本例的基础估计值如下：选择临床观察的患者有 11% 的可能发生脑卒中，而选择做手术者有 1% 的围手术期死亡，以及有 2% 发生围手术期脑卒中，术后"不发生脑卒中"者将可能有 5% 再次发生脑卒中。

如图 14-12 所示，术后患者有三种结局，即维持无症状性颈动脉硬化状态、发生脑卒中和死亡。同时随着时间的推移，无症状性颈动脉硬化状态的患者将会有一部分人死亡，一部分发生脑卒中，余下的人维持无症状性颈动脉硬化状态。对于带病生存（伴有脑卒中）者，也会随着时间推移，部分人会死亡，而生存者继续这种循环，直至所有人都死亡。

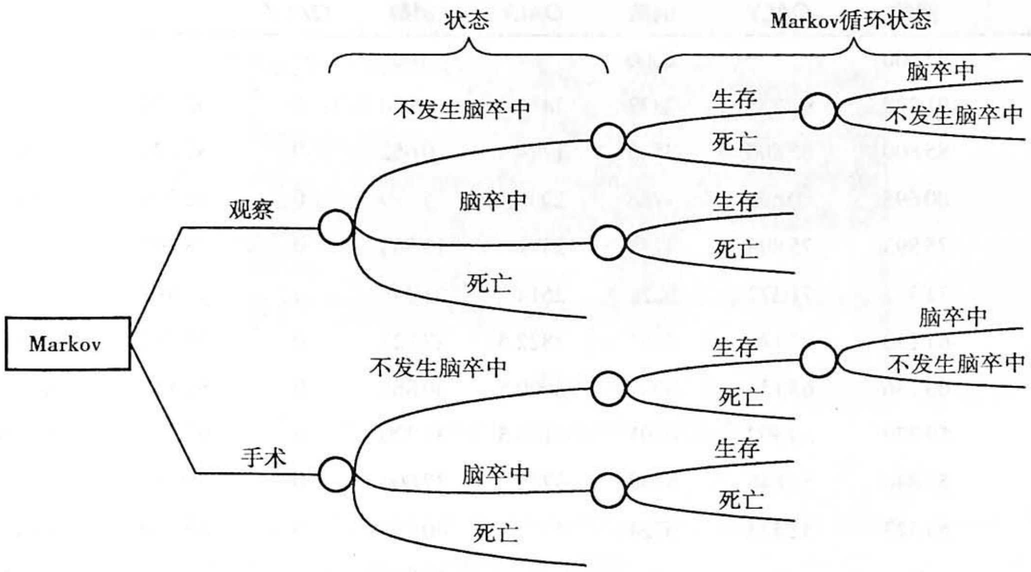


图 14-12 颈动脉硬化患者的 Markov 模型示意图（Markov 树）

用 Markov 模型来模拟他们的结局，并同其他的方案进行比较后做出选择。本例有三个 Markov 状态：无症状性颈动脉硬化（状态 1），伴脑卒中生存（状态 2），死亡（状态 3）。其中状态 1 到状态 3 的死亡率为 5%，并不完全归结于颈动脉硬化这一疾病所致（要尽可能模拟实际情况，考虑有其他死亡原因）。它们之间的转换关系及转移概率见表 14-4 和图 14-13。

表 14-4 3 种自然状态每年的转移概率

	无症状性颈动脉硬化	生存（伴脑卒中）	死亡
无症状性颈动脉硬化	0.9405	0.0095	0.05
生存（伴脑卒中）	0	0.95	0.05

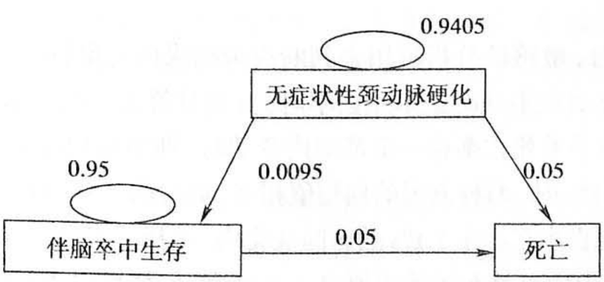


图 14-13 颈动脉硬化患者各状态间的转移关系

笔记

取 Markov 循环周期为 1 年，那么每一个周期，“无症状性颈动脉硬化”为 1 个 *QALY*，伴脑卒中生存则为 0.5 个 *QALY*，死亡则为 0。假设从手术当年起，维持无症状性颈动脉硬化占 97%，伴脑卒中生存的占 2%，死亡为 1%。此后人群在这 3 个状态的转移概率见表 14-4。

计算原理同上，鉴于这一过程的计算比较麻烦，可以借助相关软件进行，如 decision maker，treeage 等，比较简单的计算也可运用 Excel，如本例计算结果见表 14-5，对该人群而言，选择手术治疗，其最终结果是人均获得 16 个 *QALYs*。同理，计算选择“临床观察”方案的人均 *QALY*，数值较大者的方案为优选方案。

表 14-5 手术决策枝的 Markov 模型计算结果

	无症状性颈动脉硬化		生存（伴脑卒中）		死亡		总 *QALY*	累计 *QALY*
	例数	*QALY*	例数	*QALY*	例数	*QALY*		
1	97 000		2000		1000			
2	91 228	91 228	2822	1411	5950	0	92 639	92 639
3	85 800	85 800	3548	1774	10 652	0	87 574	180 213
4	80 695	80 695	4786	2093	15 119	0	82 788	263 001
5	75 893	75 893	4744	2372	19 363	0	78 265	341 266
6	71 377	71 377	5228	2614	23 395	0	73 991	415 257
7	67 130	67 130	5645	2822.5	27 225	0	69 953	485 210
8	63 136	63 136	6001	3000.5	30 863	0	99 137	551 347
9	59 379	59 397	6301	3150.5	34 320	0	62 530	613 887
10	55 846	55 846	6550	3275	37 604	0	59 121	672 998
11	55 523	52 523	6754	3377	40 723	0	55 900	728 898
50	0	0	0	0	10 000	0	0	1 599 600
合计								1 599 600
人均								16.00

注：①无症状性颈动脉硬化状态效用值 = 1，生存（伴脑卒中）效用值 = 0.5，死亡效用值 = 0；*QALY*：质量调整寿命年

②本例数据引自文献：Birkmeyer JD, Liu JY. Decision analysis models: Opening the black box [J]. Surgery, 2003, 133(1): 1-4

当然，如果对来源基础数据的可靠性存在疑问，同样可以进行敏感性分析，以了解当前决策的稳健性如何。

第五节 敏感性分析

在第四节中提到，如果认为选择做手术的死亡率（1%）过高或是过低，或者 1% 概率估计不准确，抑或不固定、是一个区间值，手术死亡率不同，将影响到最终的决策。此时可考虑进行敏感性分析。

当相关数值变化时，敏感性分析可用来判断决策结果的稳健性。这里就对第四节中的例 14-3 进行敏感性分析，当取不同的手术死亡率时，分别计算出不同的手术治疗方案的期望值，图 14-14 显示的就是当手术死亡率在一定范围内变动时、期望值相应变化情况。从图中可以看出，当手术死亡率为 2.1% 时，两种方案的期望值相当，此时的 2.1% 称为阈值，代表两种方案选择的分界点。当手术死亡率在小于 2.1% 的区间变化内，选择手术方案是合适的。

再如，对伴有脑卒中生存状态的效用值也存在疑惑，同样可以进行阈值分析和敏感性分析（图 14-15）。

笔记

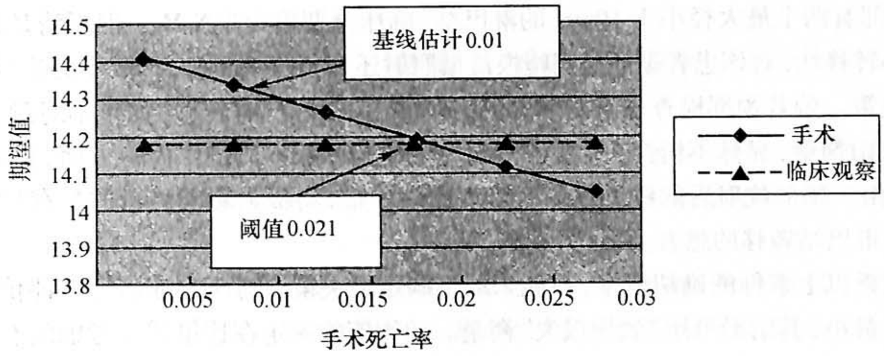


图 14-14　对手术死亡率的敏感性分析

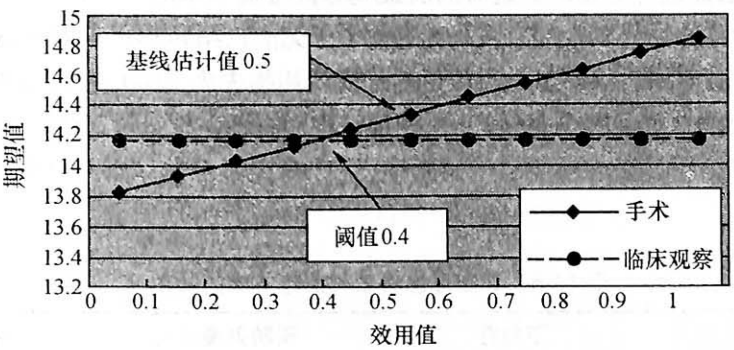


图 14-15　对伴脑卒中生存效用值的敏感性分析

实际上对事件结局及其发生概率都可以进行敏感性分析。决策树模型如此，Markov 模型也同样可以进行敏感性分析，计算出相关的阈值，方法同上。

第六节　不确定型决策

假如临床事件结局及其发生概率明确并可量化，可选择决策树模型或 Markov 模型进行决策分析。但倘若有些临床事件结局及其概率不清楚且无最佳证据可循时，只能进行不确定型决策，即依赖于决策者的临床经验和主观判断进行决策。以下准则可供决策者参考：

（一）乐观准则

也称大中取大法，即找出每种方案自然状态下的最大收益者，从中再取最大者。

（二）悲观准则

也称小中取大法，即先找出每种方案自然状态下的最小收益者，从中再取最大者。

（三）后悔值准则

计算各方案在各种自然状态下的后悔值并列出后悔值表，找出每一方案在各种自然状态下后悔值的最大者，从中再取最小者所对应的方案，即为合理方案。后悔值法利用后悔值的计算作为最优方案选择标准。先将在不同自然状态下的最大收益值作为理想目标，再把各方案的收益值与这个最大收益值的差值，作为未达到理想目标的后悔值，然后从各方案最大后悔值中取最小者，由此确定选择方案。

例 14-4　鼻咽癌以放射治疗为主，对于颈部有淋巴结转移的患者，颈部应给予根治量照射，而颈部没有淋巴结转移者，常规给予预防剂量照射，但有人也提出如鼻咽肿瘤不大，可不对颈部进行放射治疗。因此，对于颈部未发现有淋巴结转移的患者，是否给予放射治疗或给予什么剂量，目前仍有争论。假如有一鼻咽癌患者，经临床检查未发现颈部淋巴结转移和远处转移，

MR 显示颈部有两个最大径小于 10mm 的淋巴结，临床分期仍为 $T_2N_0M_0$。但不清楚患者颈部是否存在微小转移灶，对该患者颈部是实施根治量照射还是预防量照射，抑或不必进行放射治疗，需要临床决策。倘若颈部检查虽未发现淋巴结转移、但实际已有转移，如果不对颈部放疗或者只是给予预防剂量，显然不恰当，但这种情况发生的概率并不清楚。另一方面，倘若都采取根治量照射，由于颈部放射后的纤维化等后遗症比较严重，将极大影响患者的生存质量，特别是对颈部并无淋巴结转移的患者，弊远大于利。

由于缺乏以上事件的确切概率，只能采取不确定型决策。对于癌症治疗的评价，主要是使患者的损失最小，其结局可用“效用损失”衡量。效用值的确定在这里需要考虑的有：①颈部淋巴结复发或远处转移的影响；②放疗后遗症对生存质量影响的程度；③治疗费用；④患者对治疗的期望值以及放射治疗的了解等情况。同时也应考虑到：不同年龄、职业、社会地位的患者，其效用值损失的估计也不尽相同。这时的决策应尽量做到个体化。

表 14-6 中为几种结局的效用损失的定量估计：无淋巴结转移也不对颈部进行放疗的效用损失为 0，有颈部淋巴结转移而未进行放射治疗的效用损失为 −100，有颈部淋巴结转移而只进行预防剂量放射治疗的效用损失为 −30，无颈部转移而行预防剂量放射治疗的效用损失为 −50，有颈部转移而行根治剂量放射治疗的效用损失为 −20，无颈部转移而行根治剂量放射治疗的效用损失为 −80。

表 14-6 鼻咽癌患者三种治疗方案的效用值

	不放疗	预防剂量放疗	根治剂量放疗
有颈部转移	−100	−30	−20
无颈部转移	0	−50	−80

由于不知道颈部淋巴结转移的概率，面对这一不确定性，依照前面提到的几种准则进行决策：

1. **乐观准则** 在每一备择方案中，找出各种结局下的最大受益值，从中选择收益值最大的方案，也就是从最大收益中取最大。本例中不放疗的最大收益为 0，预防剂量放疗为 −30，根治剂量放疗为 −20。按照这一原则，应该选择不放疗，这是一种相对乐观的准则。

2. **悲观准则** 与乐观者相反，找出各种结局下的最小收益值，再从中选择收益值最大的方案，也就是从最小收益值中取最大。本例中不放疗的最小收益为 −100，预防剂量放疗为 −50，根治剂量放疗为 −80。最大者为 −50，应选择预防剂量放疗。

3. **后悔值准则** 按照这一准则，两种自然状态下的理想值在有颈部转移时为 −20，无颈部转移为 0，可以计算各状态下的后悔值，具体表 14-7 的后悔值矩阵。

表 14-7 后悔值矩阵

治疗方案	自然状态		最大后悔值（各方案）
	有淋巴结转移	无淋巴结转移	
不放疗	80[(−20)−(−100)]	0(0，理想值)	80
预防剂量放疗	10[(−20)−(−30)]	50[0−(−50)]	50
根治剂量放疗	0(−20，理想值)	80[0−(−80)]	80

可见在这一准则下应该选择预防剂量的放射治疗这一方案。

应该注意到，由于缺乏准确信息，利用上述决策准则时，往往带有很大的主观性。因此，决策准则及决策者不同，所得到的结果也不完全相同。当然，最好还是收集资料信息，尽量获得各种结局的概率，以改进不确定型决策，从而使选择的方案更趋合理。

笔记

第七节　决策分析的局限性

决策分析作为一种以证据为基础的临床应用技术，已开始逐步应用于临床实践中。但应注意，决策分析也存在一定的局限性，并不能保证每次都能作出正确的选择。以下几方面均可能影响决策结果：

首先，决策正确与否在很大程度上取决于数据是否充分、准确。对于决策分析所必需的基线数据，如事件概率和效用值，目前往往缺乏高质量的临床试验的结果，内部和外部偏倚会影响数据的准确性，由此可能造成决策失误。

其次，再复杂的决策模型也不能完全模拟临床实际，只是临床实际的简化版。患者病程进展往往复杂而变化无常，决策者必须考虑临床问题的方方面面，包括治疗方案、不良反应、成本，以及患者的个体差异、特殊要求等，决策中的影响因素之多远非一般决策树所能描绘。

最后，对决策分析的结果需慎重解释。决策分析的期望值为 0.96，并不能保证每个患者的实际效应值等于 0.96。它只是一个平均水平，即一组病情程度相似的患者，若选择颈部按摩后，其平均效应值为 0.96。因此，在临床实践中，应审视你的患者与决策分析中的研究对象之间的相似性，也可借助敏感性分析，为个体患者选择最恰当的治疗方案。

尽管如此，临床决策分析仍不失为一种有价值的临床实践工具，同时通过识别哪些变量（概率、效用值）是影响决策的敏感因素，决策分析也为下一步研究指明方向。

（杜进林　华贻军　洪明晃）

笔记

第十五章　循证医学自我评价

循证实践共五个步骤，其前四步，即提出问题、寻找证据、严格评价证据和应用证据于临床和预防实践，循证实践的第五个步骤亦就是循证医学自我评价，是对循证医学实践过程的评价，评价贯穿于循证实践的每一个过程。在循证医学实践过程中，发挥主导作用的是临床医生，那么医生有没有能力实践循证医学需要评价。同时实践循证医学所达到的效果如何也应评价。循证医学自我评价的方式有自我评价和同行评价。由于本教材的对象是医学生，是未来的医生，因此，本章主要针对循证自我设计、行动及效果的评价。下面将分步骤对需进行自我评价的部分进行详细介绍。

评价循证医学在某个地区、区域或国家的应用状况是卫生行政管理者、决策者所关注的内容，本章将不作讨论。

第一节　能 力 评 价

一、评价“提出可回答问题”的能力

“提出可回答问题”是循证医学实践的第一步，若问题提出的不恰当，循证医学的后续实施步骤均会受到影响。表15-1列出了5个有关“提出可回答问题”能力的自我评价问题。首先是有没有提出问题？若没有提出问题，就根本谈不上实践循证医学。其次，提出的问题其陈述是否简洁明了，符合格式？有关背景问题的两个基本成分和核心问题的三到四个成分是否准确明了。接下来的问题是，在积累了一定的经验之后，可否进一步明确主要问题所在，明确自己还欠缺哪些知识，并对最初提出的问题进行修改？另一个问题是提出问题的过程中若遇到障碍，能否想办法有能力克服之？最后一个问题是医疗实践过程中随时可能产生新的问题，有没有养成随时记录问题以待今后解决的习惯？是否将提出问题融入每天的临床活动中。记录问题的方法有多种，有沿用至今的传统方式如随身携带记事本，也有科技发展提供的高科技手段如PDA（personal digital assistant，掌上电脑）。另外，随着循证医学实践活动的增加，可查看“提出可回答问题”的成功率是否在提高，据此，可分析判断提出问题的能力是否在提升。

表15-1　对“提出可回答问题”的自我评价

条目	具体问题
1	有没有提出临床问题？
2	问题的陈述是否简洁明了，符合一定的格式？
2a	有关“背景”知识的两个基本成分
2b	有关诊断、治疗、管理等“核心”知识的三（或四）部分
3	有没有借助绘图法来明确自己还欠缺哪些知识，并对最初提出问题进行修改？
4	提出问题的过程中若遇见障碍，能否想办法克服？
5	有没有养成随时记录问题以待今后解决的习惯？

二、评价“寻找最佳外部证据”的能力

对“寻找最佳外部证据”的能力可根据表15-2中所列条目进行自我评价。首先是有没有去

笔记

寻找证据？其次，是否了解本领域内现有的最佳临床证据来源？若不知道最佳证据的来源，则很难能找到最佳证据。同时是否能以找到最高级别证据为检索目的，并力求找到位于“信息金字塔”顶端的证据（图 6-1）？第三点是能不能迅速寻找到临床实践所需的硬件、软件及最佳证据？从事循证实践人员、医院和相关其他机构是否将硬件、软件等资源合理配置？是否突出标记了最重要的证据来源？按照研究步骤如：确定资源、启动资源、键入问题、获得答案等进行时间分配时，可通过提高哪些环节的效率进而提高研究的整体效率？如果是独立开展研究，能否从庞杂的信息来源中寻找到最佳外部证据？研究能否更具效率？检索 MEDLINE 数据库时是否使用了医学主题词表（MeSH）、限制词以及智能检索（intelligent free text）等检索技巧？是否设置了有效的检索过滤器？

表 15-2　对“寻找最佳外部证据”的自我评价

条目	具体问题
1	有没有寻找证据？
2	是否了解本领域内现有的最佳临床证据来源？
3	能否迅速寻找到临床实践所需的硬件、软件及最佳证据？
4	能否从庞杂的信息来源中寻找到有用的外部证据？
5	在寻找证据的实践过程中有没有逐步提高检索的效率？
6	在 MEDLINE 数据库中检索时有没有使用截断、布尔逻辑算符（or 等）、MeSH 主题词、限制词及智能检索等检索技巧？
7	与专业文献检索人员以及热衷提供最佳证据并受人们尊敬的同行检索结果相比，我们自己的检索结果如何？

评价“寻找最佳外部证据”能力的另一个可行的方法是：请文献检索专业人员或其他专业人士对同一个问题进行重复检索，然后与我们的检索结果进行对比，比较两次检索在检索策略、检索结果、文献相关性及可用性等方面的差异。这样做有三个益处：一是能够完成自我评价；二是能从中学到更好的检索技巧；第三点是能够获得更多的、可以回答临床问题的外部证据。

除上述两种方法之外，还可以与附近的医学图书馆取得联系，参加他们举办的课程（如医学文献检索）或是索取有关的指南以便从中自我学习，达到实践循证医学所需要的文献检索能力。甚至还可以邀请文献管理员加入我们的研究团队，会更好地提高检索水平和检索效率。与循证医学有关的专业检索培训也可以有效提高文献的检索能力。

三、评估“严格评价证据质量”的能力

该项能力的评估可参照表 15-3 中的条目。首先是否对收集到的临床外部证据进行了质量评估？如果回答是否定的，那么原因是什么？能否找到并克服妨碍我们对临床外部证据进行质量评估的障碍？目前，在加拿大、美国等一些发达国家，临床医生若能加入某个小组（如各种杂志俱乐部）中，不但能有助于循证医学实践，同时还能及时获得循证实践的反馈意见。

表 15-3　对“严格评价证据质量”的自我评价

条目	具体问题
1	是否对外部证据进行了严格评估？
2	严格评估证据的指南是否易于使用？
3	能否逐渐做到准确且熟练地使用某些严格评估的指标，如似然比、*NNT* 等？
4	有没有创建任何严格评价证据的题目（CAT），对严格评估进行总结？

笔记

一旦开始评价，应考虑是否这些评估证据的指南是否容易应用。许多临床医生发现，随着时间的推移，他们能够从容评价大多数证据，但总有那么一两篇文献难以定论。这种情况在由临床医生组成的专业小组中也常会出现，但对于专业小组，问题提出的快，解决的速度也比较快。进一步需要考虑的是在严格质量评估时，对某些效应指标（如似然比、*NNT* 等）的运用是否熟练而准确。将我们的评估结果与进行同样工作的其他专业工作组的评估结果进行比较，或从二次出版期刊所摘要的文献（6S 类证据）中提取原始数据，完成数据计算，然后将计算所得结论与摘要中的结论进行验证比较。

最后，也是最高层次的要求，是否创建了任何严格评价证据的题目（critically appraised topics，CAT）并加以总结。针对严格评价的证据进行总结评分，也可用牛津循证医学中心提供的软件 CATMaker 来完成。若用 CATMaker，粗略的计算结果是否与 CATMaker 软件对文献质量评估进行总结的结果一致。虽然 CATMaker 是很有用的评价工具，实际临床工作中使用较为麻烦，也可采用更为简便的方式来评价，包括研究引用、临床底线、研究方法的两行式描述以及结果总结简表等来简要记录评估结果。

四、评价“整合外部证据与病人价值”的能力

表 15-4 中列出了一些在自我评价“结合外部证据与自我临床经验”以及评价“将检索结果应用于临床实践”时需要考虑的因素。同样，首先需要问有没有将严格评价的证据与临床实践相结合。如果没有把已经严格评估的证据应用到实践中去，那我们在前面三个步骤的努力将是白费工夫。因此，若这一步骤没有完全实施，我们需要认真反思，时间和精力将相应调整。再次强调，与良师益友进行讨论或是参与专业小组工作将非常有助于我们克服这个“纸上谈兵”的问题。工作步入正轨之后，则需反思和总结：在将外部证据应用到个体病例的过程中我们能不能做得更加准确和高效？我们是否寻找到，或是总结出了外部证据用于临床实践的验前概率？（具体可参见本教材第九章）。

表 15-4 对“外部证据与临床经验的结合”的自我评价

条目	具体问题
1	是否将严格评价的证据应用到了临床实际中？
2	能否逐渐做到准确且熟练地调整严格评价的指标（例如，验前概率，*NNT/f* 等）以适应具体的病例个案？
3	能否解释（和解决）整合证据制定决策过程中出现的争议？

我们是否能更加熟练地根据病人的“因素”（*f*）对治疗措施进行调整？能否对原有的争议进行解释（甚至是解决），是检验我们的整合能力是否有所提高的方法之一，这一评价方法尤其适用于教学医院的住院医生。

第二节 效果评价

目前，对循证医学实践的评价不仅仅是对知识技能的评价，更重要的还应包括对行为、态度的评价和对临床实践能力及临床结局改善的评价。

一、临床实践是否得到了改善

自我评价之后，了解到自我开展循证实践的能力。在具备能力的基础上，进一步应评价是否将所具有的能力转化为开展循证实践的行动，进而使临床实践得以改善。因此，应进一步自问学到的知识有没有转化为更好的临床实践？（表 15-5）。这方面可以借鉴在质量促进中，已被

笔记

证实为有效的方法，即"计划—实施—检查—行动"循环法，来帮助我们回答能力是否转化为行动的问题。其中"计划"(plan)以提出问题、寻找证据、严格评价及证据应用等4个部分作为开始。同样需要考虑：运用此方法时可能会遇到哪些障碍，以及是否需要新的技能、设备、组织程序或提醒机制来防止遗忘而没有开展行动。

表 15-5　对"行为改变"的自我评价

条目	具体问题
1	当有外部证据表明需要进行改变时，能否克服障碍进行相应的行为调整?
2	有没有进行检查，如对诊断、治疗及其他临床实践方面的审计?

在拟订要开展哪些工作的计划(plan)后，需要将计划付诸行动，亦就是实施(do)计划。计划试行一段时间后，可以看计划是否产生效果。然后检查(check)是否行动实施后有相应的效果变化。若试行后证明计划有效，则可推广，否则，需重新制订计划。检查包括对临床工作进行审计(临床审计为"一个质量改进的过程，旨在通过对卫生保健措施系统评价而对明确的标准实施变革，以改善病人卫生保健状况和疾病的结局")。需要开展审计的原因有两个：一是审计结果可以显示出临床医生的绩效如何；其二，也是更重要的一点，审计通常会以多种方式进行，特别是个体反馈，这对改善临床工作会有极大的帮助。审计工作长期、持续、恰当地进行，会不断地改善临床实践。若审计结果显示临床工作有所改善，可以进一步考虑如何继续提高自身的水平。若审计显示临床工作停滞不前，除了自我批评之外，我们还必须找到问题的根源并克服之，然后制定新的计划，进入下一循环。

二、多少临床实践有证可循

目前，许多临床工作者或专业小组研究现有的临床实践到底在多大程度上是有证可循的。实际从事临床实践的工作者也需要和有必要了解。1989年英国著名的流行病学家 Iain Chalmers 发表了他的系统综述结果：当时在产科使用的226种方法中，临床试验或系统综述证明，只有20%是有效的且疗效大于不良反应；30%是有害或疗效可疑；而50%缺乏随机试验证据。该研究结果公布后震惊全球，继而引发了对循证医学的思考和关注。

1995年由 David Sackett 领衔的审计小组，对牛津某临床机构某个月中的每一例死亡病例、出/入院病例的主要诊断(包括：疾病、症状、整体情况、存在多项诊断时导致病人就医的主诉症状)以及对病人实施的主要干预(包括：针对主要诊断所采取的治疗、操作及辅助治疗等)进行讨论，随后对干预进行追踪。审计小组基于当时的循证医学即时资源汇编及检索其他资源(如医学教科书、医学电子数据库、已发表的文献等)，最后将干预措施分为三类：

1. 已有一项或多项随机对照试验或有系统综述支持其有效性(或证明其无效)的干预；

2. 有效性已得到公认，审计小组认为再对其疗效进行随机对照试验是毫无必要的，且若设立安慰剂对照，不符合伦理学的干预；

3. 已得到普遍使用，但不属于上述两种情况者。

在该月诊断的109名患者中，根据上述的分类，90名(82%)患者接受的干预是有证可循的，其中53%的患者所接受的主要干预措施得到了一项或多项随机试验或系统评价的支持，29%的患者所接受的干预措施有明确可信的、非试验性研究的证据支持。另有18%的患者，他们接受的对症治疗或支持疗法虽无有力的证据支持，但目前的干预优于其他干预或没有干预。这次审计的结果确认了住院病人所接受的干预应该是有证可循的。从此以后，世界范围内的多个临床机构都进行了类似研究，审计涉及了临床医学的不同领域，如：普外科、血液科、儿科、基础保健、麻醉科及精神科。事实上，我们所遇见的绝大多数病人都只患有一种或两种常见病，而为常见问题寻找证据将会比为罕见问题寻找证据更加有效。上述研究还为各种常见问题累积了

笔记

众多证据，从而保证开展循证医学实践的切实可行。

最后，需要强调的是：审计不仅仅是关注如何为某个临床问题寻找到高质量的证据，审计同时更是一个逐步学习的过程，它能帮助工作小组里的每一个成员进行知识更新。

第三节　案　例

皮质类固醇产前应用于有早产风险的妇女加速胎儿肺成熟的循证实践及其效果：

早产儿是婴儿死亡的主要原因(65%)，而肺脏发育不成熟致新生儿呼吸窘迫综合征是早产儿死亡的主要原因。1972 年 Liggins 等在动物研究的基础上，首次开展了皮质类固醇产前应用于有早产风险的妇女预防新生儿呼吸窘迫综合征的随机对照试验(RCT)，至 1990 年又有 7 项 RCT 发表。然而，其效果不尽一致，1990 年 Patricia Crowley、Iain Chalmers 等首次对产前用皮质类固醇的效果开展了系统综述和 meta 分析。

该分析表明，皮质类固醇总体上以及在所分析的所有亚组均能够降低新生儿呼吸窘迫综合征的发生。皮质类固醇可降低 30%~50% 有早产风险的妇女新生儿死于早产综合征的几率。而新生儿呼吸窘迫综合征的降低，与降低脑室内出血(intraventricular haemorrhage，IVH)、坏死性小肠结肠炎(necrotizing enterocolitis，NEC)和新生儿死亡有关，同时没有强有力的证据表明皮质类固醇有副作用。

在该系统综述和 meta 分析发表之前，大多数妇产科医生没有意识到该治疗方法这样有效，而且该方法也没得到广泛使用。结果成千上万的早产儿可能罹患新生儿呼吸窘迫综合征且发生不必要的死亡(且须比实际需要更贵的治疗)。1991 年第一个 Cochrane 系统评价发表后，美国国立卫生研究院(NIH)发表共识，毫不含糊地推荐除少数以外，产前皮质类固醇治疗早产风险的妇女会大幅降低新生儿发病率和死亡率并可以节省大量的医疗费用。

在该系统综述发表前后收集的资料表明，综述发表时大约 20% 的早产风险的妇女使用皮质类固醇。该综述发表后皮质类固醇的使用在 NIH 推荐之前开始稳步增长，到 1996 年使用率攀升到 65%；同期 1991—1996 年间的资料表明，随着皮质类固醇使用的普及，婴儿死亡率相应地降低。

（赵亚双）

第十六章　临床医学研究项目申报书的撰写

科学研究的本质是创造和生产新知识，因此科研工作本身是一种非常复杂的、难度较高的、以脑力劳动为主的活动，它具有继承性、创造性、探索性等基本特点。作为一门理论发展快且实践性强的学科，临床医学的科学研究对推动临床医学整个学科发展的作用越来越显著。如何做好临床医学的科研，提升在疾病诊治、预后、病因学等方面的临床医学研究水平，是每一位临床医生和临床医学生所面临的共同挑战。因此，在了解临床医学科学研究的种类及研究基本程序的基础上，掌握临床医学研究项目申报书要领、撰写出高质量的申报书，是医院学术水平和综合研究能力的具体体现，也是学科发展和临床医学工作者基本素质的标志。

第一节　概　　述

一、科学研究的基本程序

科学研究应涵盖从提出问题、验证假说到得出结论等全过程。其基本程序和具体步骤包括：查阅文献、提出临床医学科学问题和理论假说、制订研究计划及设计研究方案、撰写项目申报书、获得资助后的项目启动与实施（实验观察或调查）、研究资料的整理与数据处理、统计分析、研究结果的归纳和总结、撰写研究报告以及成果转化等。

二、临床科研的基本类型

要撰写高质量的临床医学研究项目申报书，首先要了解临床科研的具体类型。按研究任务来源、科技活动类型及研究内容可归纳为以下几方面：

1. 按任务来源可分为纵向科研任务、横向科研任务及自由选题项目。其中：①纵向科研任务是指各级政府主管部门下达的课题及项目，包括国家自然科学基金委员会设立的各类科学研究基金，政府管理部门科研基金，如科技部／卫生部的科学研究基金，以及单位科研基金。②横向科研任务是以横向科技合同为依据的项目，它主要由企业、事业单位及其他机构委托进行，研究经费一般由委托单位提供。③自由选题，是根据学科发展和科技人员的专长，并结合医疗卫生工作的实际需要，由科技人员自己提出的研究课题。

2. 按科技活动类型可分为：基础研究、应用研究及发展研究。其中：①基础研究是以认识自然现象、探索自然规律为目的，未涉及或仅笼统涉及社会应用设想的研究活动。②应用研究，主要针对某个特定的、有实际应用价值的目标开展的研究活动。③发展研究，也称开发研究，是运用基础研究和应用研究的成果，旨在推广新材料、新产品、新设计、新流程和新方法，或对之进行重大的、实质性改进的创造活动。

三、科研项目申请前的准备

（一）文献调研

文献查阅和复习，既是课题选择的前提，也是产生或形成临床医学科学问题假设的前提。通过文献查阅，可以把握本学科的整体发展水平，找出当前研究的热点和焦点问题，了解这些问题在国内外的研究现状；根据文献调研，还可预测本学科及相关学科的发展趋势以及未来的发展方向，帮助形成课题研究的基本思路。

笔记

进一步将拟研究的题目与相关研究文献进行比较，精读文献，了解验证研究假设的技术方法以及在本学科和本专业的应用现状，思考拟进行的研究可以探讨哪些临床医学问题，有哪些特色和创新之处，进一步构思研究方案。

（二）研究基础与预实验

明确相关研究基础以及与本次选题的关系，若无现成的文献资料可供参考，也可进行预实验或小规模的预调查，以进一步理清思路、找准问题、选题立题，为撰写申报书做好铺垫。

第二节 研究设计报告书的基本内容与撰写方法

一、课题名称及摘要

（一）申报课题名称

课题名称是对申请课题的研究方法、研究对象、研究目标的综合描述，要体现出特色、深度及创新，表达要简洁明了、语言修辞要正确、题目要有吸引力和可信度。

（二）申报书摘要

摘要是申报书的高度概括，有固定的字数要求，一般在400字左右，要求所描述内容充满新意，阐述的条理非常清楚。摘要应介绍所申请项目的研究背景和现状、存在问题、立题依据，研究目的和重要意义，研究方法与主要技术路线，以及预期研究结果等。

撰写摘要时，应围绕下列问题依次展开：即目前存在的主要科技或者科学问题是什么？为何要进行研究？本研究拟采用的主要方法及技术手段是什么？可以达到的研究目标及意义是什么（包括科学意义/潜在价值或者应用前景）？

初次撰写摘要时，也可参考如下撰写格式，填写相关内容："……是……防治的瓶颈，但其机制尚未明了。本课题拟采用……手段（模型、技术、方法），研究……机制（环节），特别是……的规律（作用）。应用……，进一步探讨……的新方法（技术）。为……提供科学的理论依据和安全有效地……方法，必将具有广阔的应用前景"。要特别注意摘要与申报书的所有内容一致、前后呼应。

二、立项依据

项目的立项依据是判定研究设计是否具有创新性、科学性、可行性、实用性、效能性的基础。该部分需要具体阐述研究意义、国内外研究现状及发展动态，结合科学研究发展趋势来论述科学意义或应用前景。同时应附主要参考文献目录。

（一）介绍申报项目提出的重要科学问题

本项目所关注的一定是国内外尚未解决或者亟待解决的重要科学问题，以及临床实践中发现的、具有重大意义的研究问题。特别要突出研究对国民经济和社会发展的重要性、迫切性。

（二）描述国内外相关问题的研究现状及发展动态

重点阐述国内外相关问题的研究现状，特别是与本课题研究思路、拟采用的技术方法相似的研究现状，注意提炼、归纳、总结出学科研究中存在问题、热点和难点；同时还要对国内外研究的发展动态进行分析，这些都是形成立题思路和提出立题点的关键，用以充分说明这些问题是亟待解决、需深入探讨的重要科学问题。

（三）申报项目的假设和研究思路

提出本次申报项目的理论设想或假设，介绍拟通过什么研究思路、采用何种方法和技术手段、重点解决哪些问题。

笔记

（四）参考文献引用的注意要点

申报书的撰写，特别是对其创新性、科学性和应用价值等的阐述和论证，要有充分的依据，这些都需要有可靠权威的文献材料来支撑（包括引用相关文献中的论题、观点、概念、理论、方法、结果、结论或重要事实、数据等）。通过追踪和引述本研究领域近期发表的相关文献，阐明本项目的前沿性或新颖性，提出创新点。近3年的文献量应该占2/3左右，最好引用一定比例的当年参考文献，以展示申请者知晓国内外最新研究进展和发展动态；同时注意引用国内外重要的、高质量的研究文献，以及引用与自己研究高度相关的参考文献，以论证或说明所申报项目的科学性和重要性。

三、研究目标、研究内容及拟解决的关键科学问题

（一）研究目标

指研究拟要达到的目的（最终目标），是对题目、研究对象、研究方法、成果和应用的高度概括和具体化。将研究内容中阶段性成果简单罗列在一起不能算作研究目标。撰写研究目标，可参考如下格式："通过……（技术和手段），解决……的问题、证实（验证、阐明）……的科学假设（假说），……环节？解决（揭示）……的关键科学（技术）问题，预期达到……目的"。

（二）研究内容

研究内容是对研究目标的深层次诠释，或者说是对研究目标的具体说明。撰写"研究内容"时，要围绕研究目标逐层、逐次展开，强调为了实现研究目标，研究分几个步骤进行，拟从何处入手，重点研究哪个方面，主攻什么方向，达到什么样的预期效果以及有哪些具体的考核和衡量指标等。

研究内容要具体适当、重点要突出、层次要分明、用词要严谨、研究任务要明晰。切忌内容过多、研究周期内难以完成；或内容分散，研究目标无法逐一实现。

（三）拟解决的关键科学问题

拟解决的关键科学问题要反映申请者对课题总体目标实现的深刻理解和统筹解决的能力。该部分主要涉及研究过程中的一些关键技术问题、科学手段等，这些关键问题决定了研究能否有科学新发现、实验方法学有新突破、技术有创新等。注意技术和设备上或者课题合作时可能出现的困难等，不能当做拟解决的关键科学问题。此部分撰写时，要陈述对关键问题拟采用的解决方案及备用解决方案等。

四、研究方案与可行性分析

研究方案是申报书中文字篇幅最大的部分，是在研究目标和研究内容确定之后，对项目具体实施方案的详尽描述，撰写方法可因人而异，以时间顺序为主线或以研究内容为主线描述研究方案均可，但要分大小标题，突出逻辑关系，详细地写清楚每个具体步骤，具体阐述研究方法和技术路线，对复杂的技术路线和研究方法可采用流程图或图表来说明。研究方案撰写时要注意叙述清晰详细和上下文逻辑性，过于简单和过于繁杂都不可取。下面分几个方面重点介绍：

（一）研究因素

一项研究在决定了研究目标和内容后，就应该确定合适的研究因素，即研究变量。描述所选择的变量（因变量和自变量）及其规定，因变量也叫结局变量，一般指发病、死亡、伤残、治愈、有效、无效、某个观察指标变化等事件。自变量就是影响或者决定因变量的有关因素，如疾病史、家族史、疾病类型、基础疾病、药物种类和剂量以及行为生活方式（饮食习惯、食物种类等生物因素，体力活动等行为因素），一项研究中涉及的自变量和因变量可能很多，大多能通过病历、专门设计的调查表和记录表等途径加以测量。在选择变量，特别是确定自变量时一定要充分考

虑到它与因变量之间的生物学关系及逻辑学上的合理性。

申报书应介绍研究变量的规定和说明，将这些规定书写成文，例如，疾病的诊断及预后标准等。

（二）研究方法

研究目的不同，研究方法也不尽相同，同一项研究也可用多种研究方法。研究方法的陈述主要包括所用的研究设计类型、生物学技术与方法、统计学分析方法、干预方法、资料收集方法等，均应说明，以体现研究设计的合理性和可行性，以及能够获得真实可靠研究结果的可能性。

1. 研究设计类型 应确定在研究中用人群研究的方法，还是进行动物实验或体外研究。人群研究的设计类型主要有观察性研究和实验性研究。可用观察性研究中的现况研究方法调查某一疾病在人群的患病率，如慢性肾病的流行病学调查；用病例对照研究和队列研究进行病因学和危险因素的研究，如中青年脑梗死危险因素的病例对照研究，妊娠期高血压对低出生体重影响的队列研究等；用实验性研究中的临床试验方法进行药物疗效的评价，如某新药治疗原发性高血压的多中心随机对照临床试验，等等。

2. 生物学技术与方法 介绍项目所用的细胞生物学技术、蛋白分析技术、核酸分析技术等。如用流式细胞仪测定活细胞的某种蛋白，用酶联免疫吸附试验测定血清中某种蛋白，用实时定量 PCR 的方法扩增核酸等。

3. 干预方法 临床医学研究中的干预措施至关重要，是决定研究成败的关键之一。干预措施种类繁多，干预方法各有不同，要做详尽介绍。如某种药物的疗效评价，干预措施为给予研究对象某种药物，要确定并详细介绍干预组和对照组用药的药物名称、生产厂家、剂量、疗程、给药途径、疗效和不良反应的观察方法、不良反应的处理措施等。当然，疗效评价研究中所用干预措施还可以是一种治疗方案和策略或其他干预措施的疗效评价。对照组种类很多，标准对照即给予常规药物，安慰剂对照需给予安慰剂，常规药物和安慰剂用药情况同样应做详细介绍。另外，干预措施的依从性是影响药物疗效的重要问题，提高依从性和依从性监测的方法要有合理的安排。

4. 资料收集方法 资料的收集方法是临床研究设计的重要内容。获取资料的方法一定要介绍清楚，一般有两种，一种是指按研究设计方案通过专门调查或检查检验所收集的资料，如现场询问、信访或电话访问、医学检验、体检或直接观察中获得；另一种是从个人记录（病历等医学记录）或某些部门的资料（人口普查、生命统计、死亡登记、疾病登记、电子健康档案）中获得。如临床经济学分析，需要说明这些费用数据的来源，直接成本和间接成本计算的依据和种类。调查表是资料收集的最主要工具之一。

绝大多数的研究都是通过调查表来收集所需的资料或信息的。在申请书后要附所用调查表，并且应介绍保证调查真实性的具体措施。

要介绍生物学材料，如研究对象的血、尿、便或组织标本等的采集方法，储存条件，以及用采集的生物学标本进行何种实验室检验或其他检查。以及所用仪器设备的型号、生产厂家等。

5. 统计学分析方法 申报书中要介绍数据分析的方法，要列出分析的主要项目及指标、所应用的统计学方法，如治疗组与对照组血压值的比较方法，有效率及治愈率的比较方法，多因素模型的建立方法，对建模因素的选择、混杂因素的处理以及因素间的交互作用分析等。要介绍统计学分析方法的名称，如卡方检验、相对危险度及其 95% 可信区间、logistic 回归模型、Cox 比例风险模型等，具体何种研究采用何种分析方法，请参见有关医学统计学书籍的详细论述。

（三）研究对象及样本大小

根据研究目的、研究内容、研究方法选择研究对象，介绍选择研究对象的依据，研究对象来源、纳入及排除标准、随机抽样和分组的方法、样本大小计算的依据和方法，等等。

1. 来源 研究对象的来源要根据研究的性质和方法来确定。如果研究对象为病人，可从医

笔记

院、门诊的病人中选取，也可从社区的疾病监测资料或普查、抽查的人群资料中获得，有时也可利用单位团体中所有被诊断的病例。一般以社区人群来源最优，它的代表性强，但不容易得到，故这种情况多采用医疗机构的病人。如果同时要为病人选取对照，必须与病人来源于一个总体。

2. 选择标准　即根据什么标准选择研究对象。如可以限定研究对象的人口学特征，包括性别、年龄、民族、职业等；如果研究对象是某一种病人，要说明该病的诊断标准及依据、疾病的类型等；为防止研究因素以外因素的影响，最好剔除具有所研究疾病或健康状态的另外一些影响因素的对象，如存在与研究因素有联系的其他疾病或健康状态时，亦应予以剔除；为提高研究效率，可以将研究对象限制在居住于一定地区及一定年限。

3. 随机抽样和随机分组方法　大多数研究是不可能以目标人群的全体作为研究对象的，这里就涉及一个抽样的问题。一项研究采用什么样的抽样方法应在设计中交代清楚，有助于估计研究的可靠性和精确性。如，在流行病学调查研究中使用的抽样方法主要有单纯随机抽样、系统抽样、分层抽样、整群抽样和多级抽样。在临床试验中如何进行随机分组，等等。具体见本教材第三章有关内容。

4. 样本大小估计　在流行病学研究中，增加样本含量是减少抽样误差、提高研究精确性的最基本的方法。但样本过大会徒然浪费人力、物力、时间和费用，因此，一般要根据研究目的与需要确定合适的样本含量。样本含量的估计一般根据研究的类型确定，应介绍所用统计学中样本大小的计算公式以及估计依据。

（四）技术路线

技术路线是项目实施的流程图，是针对研究目标、研究内容、研究方法等所制定的合理可行的路径。用图文混排的形式表示，可能更为清晰、简单，一目了然。撰写时一定要与项目的研究目的、研究内容、研究目标等环节相互衔接，达到研究目标明确、研究内容充实、研究范围合理、实验设计较为完善、实施过程可行的目的。

（五）可行性分析

可行性分析主要考察课题组是否有能力和条件完成所申请的项目，即已经具备的人员、设备、技术或者方法学上的能力和保证，内容包括：

1. 条件及资源　已具备项目所需要的实验环境条件、设备与物资，以及可以从医院或社区收集到的研究所需的病例及其他研究对象，所用的实验动物等。

2. 工作积累　课题组前期的相关工作积累已经为所申请项目创造了条件。

3. 研究团队能力　人员结构合理，如研究团队中有教授、副教授，或主任医师及副主任医师作指导，掌握关键技术的技术人员，研究团队成员的前期工作或研究背景与本次申请项目有关，能够胜任申请项目所承担的工作。具备了完成课题的能力。

（六）质量控制

一项临床医学研究的目的是为了展示临床医学中的客观现象，揭示其真实规律，任何研究者均希望将误差或偏倚控制到尽可能小，研究质量的控制是贯穿于整个研究过程中的，因此，应根据相关专业知识分析在整个研究过程中每个阶段产生误差或偏倚的可能性，如研究对象的选择过程、资料收集阶段、资料整理、资料分析等环节，制定正确的质量控制的对策或措施，有关在研究各环节可能出现的误差或偏倚的来源及其测量与控制方法，有关章节已有详细介绍，在制定质量控制对策与措施时，可根据不同的研究目的与研究方法参照有关内容，制定具体的切合实际的质量控制方法。

五、项目的特色与创新之处

笔记

特色与创新即在本项目研究领域中申请者与国内外同行所不同。应从对包括项目的立论

依据、研究内容、研究方法与手段、技术路线及其实验方案上的研究与创新点进行概括、提炼并集中反映出来。主要包括理论的创新，技术或方法的创新。可以做如下描述："通过……实验、临床过程，提出……研究的假设，通过……手段和方法证明……；采用……实验方法，实验技术解决……中……关键问题；本研究的特色是从……方面入手，对……的问题进行研究，期望阐明……，揭示……机制。这些研究在国内外鲜有报告。此外，通过……的研究，为临床的应用提供……依据"。

六、年度研究计划及预期研究结果

（一）年度研究计划

合理地划分年度计划进展，体现所制定的目标能够完成。此部分撰写的目标基本上是按照课题的研究内容、研究方法和技术路线所制订的方案进行的。一般为准备或预实验预调查阶段，项目实施阶段，项目总结阶段。各阶段完成的内容，具体实施计划，以及不同阶段之间的衔接要做清晰的介绍。

（二）预期研究成果

预期研究成果撰写要涵盖理论成果，应用成果等。预期研究成果与"研究特色和创新"基本上应该一致。即在某一理论方面有所建树或在某项关键技术、方法学上获得创新。

预期研究成果中应报告论文撰写的篇数和质量，包括国内外期刊、国内外学术会议论文等。也包括：拟组织或参加的国内外重要学术交流活动、国际合作项目、交流计划、专利等。

七、研究基础与研究团队介绍

（一）研究基础

研究基础撰写时要介绍申请项目前期的工作积累。注意描述通过前期工作，所发现的与本次研究相关的问题与线索，前期工作成绩为本次申请项目奠定了基础，如：建立的某一方法学技术，建立的某种疾病预后研究的队列等。工作基础不仅仅指申报者个人的工作基础，要特别注意整个研究团队、学科或者学校的相关工作基础、人员和设备的资源。

（二）研究团队介绍

研究团队包括申请人及项目组成员。介绍内容包括申请人情况及主要成员情况。如申请人的学历、职称、从事研究和科研训练经历、主要研究方向，与本题相关的研究工作和课题（特别要说明本次申请项目是否为既往参加或主持的研究工作的延续），以及与项目有关或者相关的获奖和成果，代表性论文与级别等。主要成员一般为6~10人，年龄、职称、人员结构搭配要合理，一般有高级研究人员1~2人，中级研究人员2~3人，技术人员及研究生3~5人。各自所从事的工作衔接良好。描述要切题，与此项目无关的内容要删除。

八、医学伦理学问题

医学伦理学问题是进行医学科学研究必须考虑的，应严格遵守《世界医学协会赫尔辛基宣言》（简称《赫尔辛基宣言》）和《人体生物医学研究国际道德指南》制定的涉及人体对象医学研究的道德原则，包括以人作为受试对象的生物医学研究的伦理原则和限制条件，人体生物医学研究政策和伦理标准以及已确立的伦理审查机制等。

九、经费预算

科研经费是进行科研活动的基本保障。应按照相关规定，本着实事求是、精打细算的原则，编制切合实际的项目经费预算。项目资助经费预算包括收入预算与支出预算。收入预算包括用于项目研究的各种不同渠道的经费，如从自然科学基金获得的资助、从项目依托单位获得的

笔记

资助和从其他渠道获得的资助等。支出预算包括研究经费、国际合作与交流经费、劳务费和管理费。在项目申请时，支出预算要详尽合理列出，一般依据以下几个方面进行预算：

（一）研究经费

研究经费是指直接用于科学研究的费用，是经费预算的重点，占申请经费的80%左右，包括科研业务费、实验材料费、仪器设备费、实验室改装费、协作费等。科研业务费含测试、计算、分析费、差旅费、调研和学术会议费、论文版面费和资料印刷费，以及文献检索、入网等信息通信费等；实验材料费含原材料、试剂、药品等消耗品购置费，实验动物、标本、样品的采集加工费和包装运输费等；仪器设备费包括专用仪器设备购置、运输、安装费和修理费，自制专用仪器设备的材料、配件购置费和加工费；实验室改装费为改善资助项目研究的实验条件，对实验室进行改装所开支的费用；协作费为外单位协作承担项目部分研究试验工作的费用。

（二）国际合作与交流经费

国际合作与交流经费是指用于与资助项目研究工作有直接关系的国际合作与交流费用，包括项目组人员出访及外国专家来访的部分费用。一般小额度的资助项目无此项预算。

（三）劳务费

劳务费是指用于直接参加项目研究的学生的劳务费用。一般不得超过资助经费的10%~15%。

（四）管理费

管理费是指项目依托单位为组织和支持项目研究而支出的费用，包括项目执行中公用仪器设备、房屋占用费等。一般管理费不得超过资助经费的5%。

所有项目支出计算应该分配合理，逐一列出，写明计算依据。仪器设备费，实验室改装费尽量少填，最好不填，不然会给人造成研究条件尚不具备的印象。

十、对照检查提纲

一份项目申请书初稿撰写完毕，要反复审慎地审查、修改、再审查、再修改，项目组成员要多次讨论，直至认为基本没有什么纰漏后，可请相关专家学者审阅，提出批评建议，也可请从未涉足申请项目相关问题研究的老师和同学审阅，以评价所申报项目立论依据是否充分，研究目标是否明确，假设是否合理，研究内容是否围绕研究目标进行，方案合理可行否，项目的特色及创新性怎样，以下几点列出临床科研设计的检查提纲，可供参考：

1. 课题研究的意义是否阐释清楚，是否为重要领域的重要问题，是否具有重要的理论价值或应用前景。

2. 国内外目前的研究现状怎样，了解的广度和深度如何，近5年的主要研究进展是否掌握，是否介绍了目前研究存在的主要问题，以此提出所申请项目的切入点，进一步形成研究目标及思路。

3. 项目特色及创新性体现在哪方面，是否可能形成新的学说或理论，是否产生新的方法与技术，或改进及完善了以往的技术。

4. 尽可能引用国内外近3~5年的参考文献。

5. 研究目标是否能够实现，是否目标过大或过于局限。

6. 研究内容范围是否合适，一般列出3~5个研究内容，要有1~2个突出的重点，把握1~2个关键问题。

7. 研究方案设计是否合理，研究变量选择是否合理，研究对象纳入排除标准是否明确，样本量是否足够，方法与技术是否成熟可靠、可重复性强、易于掌握。

8. 与本项目有关的工作积累有哪些，有无已发表的相关研究论文和工作基础，预实验的结果如何以及是否建立了实验模型。

笔记

9. 开展研究的条件是否具备，主要仪器和设备是否到位，病例的积累是否足够等。

10. 项目组成员结构是否合理：主要成员组成是否合理，高、中级研究人员，技术人员以及学生的搭配比例如何。注意大学生创新项目的主要成员应为在校的大学生。

（王素萍）

笔记

附表　随机数字表

编号	1~10										11~20										21~30										31~40										41~50									
1	22	17	68	65	84	68	95	23	92	35	87	02	22	57	51	61	09	43	95	06	58	34	82	03	47
2	19	36	27	59	46	13	79	93	37	55	39	77	32	77	09	85	52	05	30	62	47	83	51	62	74
3	16	77	23	02	77	09	61	87	25	21	28	06	24	25	93	16	71	13	59	78	23	05	47	47	25
4	78	43	76	71	61	20	44	90	32	64	97	67	63	99	61	46	38	03	93	22	69	81	21	99	21
5	03	28	28	26	08	73	37	32	04	05	69	30	16	09	05	88	69	58	28	99	35	07	44	75	47
6	93	22	53	64	39	07	10	63	76	35	87	03	04	79	88	08	13	13	85	51	55	34	57	72	69
7	78	76	58	54	74	92	38	70	96	92	52	06	79	79	45	82	63	18	27	44	69	66	92	19	09
8	23	68	35	26	00	99	53	93	61	28	52	70	05	48	34	56	65	05	61	86	90	92	10	70	80
9	15	39	25	70	99	93	86	52	77	65	15	33	59	05	28	22	87	26	07	47	86	96	98	29	06
10	58	71	96	30	24	18	46	23	34	27	85	13	99	24	44	49	18	09	79	49	74	16	32	23	02
11	57	35	27	33	72	24	53	63	94	09	41	10	76	47	91	44	04	95	49	66	39	60	04	59	81
12	48	50	86	54	48	22	06	34	72	52	82	21	15	65	20	33	29	94	71	11	15	91	29	12	03
13	61	96	48	95	03	07	16	39	33	66	98	56	10	56	79	77	21	30	27	12	90	49	22	23	62
14	36	93	89	41	26	29	70	83	63	51	99	74	20	52	36	87	09	41	15	09	98	60	16	03	03
15	18	87	00	42	31	57	90	12	02	07	23	47	37	17	31	54	08	01	88	63	39	41	88	92	10
16	88	56	53	27	59	33	35	72	67	47	77	34	55	45	70	08	18	27	38	90	16	95	86	70	75
17	09	72	95	84	29	49	41	31	06	70	42	38	06	45	18	64	84	73	31	65	52	53	37	97	15
18	12	96	88	17	31	65	19	69	02	83	60	75	86	90	68	24	64	19	35	51	56	61	87	39	12
19	85	94	57	24	16	92	09	84	38	76	22	00	27	69	85	29	81	94	78	70	21	94	47	90	12
20	38	64	43	59	98	98	77	87	68	07	91	51	67	62	44	40	98	05	93	78	23	32	65	41	18
21	53	44	09	42	72	00	41	86	79	79	68	47	22	00	20	35	55	31	51	51	00	83	63	22	55
22	40	76	66	26	84	57	99	99	90	37	36	63	32	08	58	37	40	13	68	97	87	64	81	07	83
23	02	17	79	18	05	12	59	52	57	02	22	07	09	47	03	28	14	11	30	79	20	69	22	40	98
24	95	17	82	06	53	31	51	10	96	46	92	06	88	07	77	56	11	50	81	69	40	23	72	51	39
25	35	76	22	42	92	96	11	83	44	80	34	68	35	48	77	33	42	40	90	60	73	96	53	97	86
26	26	29	13	56	41	85	47	04	66	08	34	72	57	59	13	82	43	80	46	15	38	26	61	70	04
27	77	80	20	75	82	72	82	32	99	90	63	95	73	76	63	89	73	44	99	05	48	67	26	43	18
28	46	40	66	44	52	91	36	74	43	53	30	82	13	54	00	78	45	63	98	35	55	03	36	67	68
29	37	56	08	18	09	77	53	84	46	47	31	91	18	95	58	24	16	74	11	53	44	10	13	85	57
30	61	65	61	68	66	37	27	47	39	19	84	83	70	07	48	53	21	40	06	71	95	06	79	88	54

续表

编号	1~10					11~20					21~30					31~40					41~50				
31	93	43	69	64	07	34	18	04	52	35	56	27	09	24	86	61	85	53	83	45	19	90	70	99	00
32	21	96	60	12	99	11	20	99	45	18	48	13	93	55	34	18	37	79	49	90	65	97	38	20	46
33	95	20	47	97	97	27	37	83	28	71	00	06	41	41	74	45	89	09	39	84	51	67	11	52	49
34	97	86	21	78	73	10	65	81	92	59	58	76	17	14	97	04	76	62	16	17	17	95	70	45	80
35	69	92	06	34	13	59	71	74	17	32	27	55	10	24	19	23	71	82	13	74	63	52	52	01	41
36	04	31	17	21	56	38	73	99	19	87	26	72	39	27	67	53	77	57	68	93	60	61	97	22	61
37	61	06	98	03	91	87	14	77	43	96	43	00	65	98	50	45	60	33	01	07	98	99	46	50	47
38	85	93	85	86	88	72	87	08	62	40	16	06	10	89	20	23	21	34	74	97	76	38	03	29	63
39	21	74	32	47	45	73	96	07	64	52	09	65	90	77	47	25	76	16	19	33	53	05	70	53	30
40	15	69	53	82	80	79	96	23	53	10	65	39	07	16	29	45	33	02	43	70	02	87	40	41	45
41	02	89	08	04	49	20	21	14	68	86	87	63	93	95	17	11	29	01	95	80	35	14	97	35	33
42	87	18	15	89	79	85	43	01	72	73	08	61	74	51	69	89	74	39	82	15	94	51	33	41	67
43	98	83	71	94	22	59	97	50	99	52	08	52	85	08	40	87	80	61	65	31	91	51	80	32	44
44	10	08	58	21	66	72	68	49	29	31	89	85	84	46	06	59	73	19	85	23	65	09	29	75	63
45	47	90	56	10	08	88	02	84	27	83	42	29	72	23	19	66	56	45	65	79	20	71	53	20	25
46	22	85	61	68	90	49	64	92	85	44	16	40	12	89	88	50	14	49	81	06	01	82	77	45	12
47	67	80	43	79	33	12	83	11	41	16	25	58	19	68	70	77	02	54	00	52	53	43	37	15	26
48	27	62	50	96	72	79	44	61	40	15	14	53	40	65	39	27	31	58	50	28	11	39	03	34	25
49	33	78	80	87	15	38	30	06	38	21	14	47	47	07	26	54	96	87	53	32	40	36	40	96	76
50	13	13	12	66	99	47	24	49	57	74	32	25	43	62	17	10	97	11	69	84	99	63	22	32	98

摘自 Fisher RA, Yates F. Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, 1957: 128

参考文献

[1] 王家良，王滨有. 临床流行病学. 第3版. 北京：人民卫生出版社，2008

[2] 王家良. 循证医学. 第2版. 北京：人民卫生出版社，2010

[3] 王家良. 临床流行病学——临床科研设计、测量与评价. 第3版. 上海：上海科技出版社，2009

[4] 洪明晃. 临床科学研究设计、测量、评价. 第2版. 广州：中山大学出版社，2002

[5] Sackett DL, Straus SE, Richardson WS, et al. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Churchill Livingstone, 2000

[6] 李定国，黄红. 医院流行学教程. 北京：科学出版社，2002

[7] Sharon E. Straus, W. Scott Richardson, Paul Glasziou, R. Brian Hayne. 循证医学实践和教学. 詹思延，译. 北京：北京大学出版社，2006

[8] Aleem IS, Karamlou T, Benson LN, McCrindle BW. Transcatheter device versus surgical closure of ventricular septal defects: a clinical decision analysis. Catheter Cardiovasc Interv, 2006, 67: 630-636

[9] Haynes B, Sackett DL, Guyatt GH, Tugwell P. Clinical Epidemiology: How to Do Clinical Practice Research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2005

[10] 杨克虎. 循证医学. 北京：科学出版社，2007

[11] Aleem IS, Jalal H, Aleem IS, et al. Clinical decision analysis: Incorporating the evidence with patient preferences, 2009, 3: 21-24

[12] Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 1996

[13] Robert H. Fletcher, Suzanne W. Fletcher. Clinical epidemiology. 4th ed. Philadelphia: Lippincott Williams & Wilkins, 2005

[14] Steven Piantadosi. Clinical trials. New York: John Wiley & Sons, 1997

[15] Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Raven Publishers, 1998

[16] 张俊华，商洪才，张伯礼. 系统评价和meta分析质量的评价方法. 中西医结合学报，2008，6(4)：337-339

[17] 中华医学会神经病学分会脑血管病学组急性缺血性脑卒中诊治指南撰写组. 中国急性缺血性脑卒中诊治指南2010. 中华神经科杂志，2010，43(2)：146-153

[18] Geyman JP, Deyo RA, Ramsey SD. Evidence-based clinical practice: concepts and approaches. Eds: Butterworth Heinemann, Boston, 2000, 1-11, 119-123

[19] Whitney SN, McGuire AL, Laurence JD, McCullough B. A typology of shared decision making, informed consent, and simple consent. Ann Intern Med, 2003, 140: 54-59

[20] 李立明. 流行病学. 第6版. 北京：人民卫生出版社，2007

[21] 李立明. 临床流行病学(长学制). 北京：人民卫生出版社，2011

[22] Steinbrook R. Searching for the right search--reaching the medical literature. N Engl J Med, 2006, 354(1): 4-7

[23] 黄悦勤. 临床流行病学. 北京：人民卫生出版社，2002

[24] Kaplan RM. Shared decision making: a new tool for preventive medicine. Am J Prev Med, 2004, 26(1): 81-83

[25] Paul GS, steven HW, Martin E, et al. Developing Guidelines. BMJ, 1999, 318: 593-596

[26] 谭红专. 现代流行病学. 北京：人民卫生出版社，2005

[27] Guyatt G, Rennie D(EDs). Users' Guides to Medical Literature: Essentials of Evidence-based Clinical Practice. JAMA & Archives Journals, AMA Press, 2002

[28] 詹思延. 流行病学. 第7版. 北京：人民卫生出版社，2012

[29] Richardson WS, Detsky AS. Users' guides to the medical literature, VII: how to use a clinical decision analysis, B: what are the results and will they help me in caring for my patients? JAMA, 1995, 273: 1610-1613
[30] Woof SH. Practice guideline, a new reality in medicine. I. Recent development. Arch Intern Med, 1990, 150: 1811-1818
[31] 孙贵范. 预防医学. 北京:人民卫生出版社, 2006
[32] 黄悦勤. 临床流行病学(研究生教材). 第2版. 北京:人民卫生出版社, 2010
[33] Ho K, Chockalingam A, Best A, Walsh G. Technology Enabled Knowledge Translation: building a framework of collaboration. CMAJ, 2003, 168: 710-711
[34] 施侣元. 流行病学词典. 北京:科学出版社, 2001
[35] Naglie G, Krahn MD, Naimark D, et al. Primer on medical decision analysis: Part 3--Estimating probabilities and utilities. Med Decis Making, 1997, 17(2): 136-141
[36] Woof SH. Practice guideline, a new reality in medicine. Ⅱ. Methods of developing Guideline. Arch Intern Med, 1992, 152: 946-952
[37] 李幼平. 循证医学. 第2版. 北京:高等教育出版社, 2010
[38] 常存库. 揭开生命与疾病奥秘的钥匙:医学科学方法学. 北京:中国协和医科大学出版社, 2006
[39] Bero LA, Grilli R, Grimshaw JM, et al. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. BMJ, 1998, 317(7156): 465-468
[40] Krahn MD, Naglie G, Naimark D, et al. Primer on medical decision analysis: Part 4. Analyzing the model and interpreting the results. Med Decis Making, 1997, 17: 142-151
[41] 于小松. 全科医学理论与循证实践. 北京:人民卫生出版社, 2012
[42] 贺石林, 陈修. 医学科研方法学导论. 北京:人民卫生出版社, 1998
[43] John Page, Richard Heller, Scott Kinlay, et al. Attitudes of developing world physicians to where medical research is performed and reported. BMC Public Health, 2003, 3: 6
[44] 赵仲堂. 流行病学研究方法与应用. 第2版. 北京:科学出版社, 2005
[45] Naimark D, Krahn MD, Naglie G, et al. Primer on medical decision analysis: Part 5--Working with Markov processes. Med Decis Making, 1997, 17(2): 152-159
[46] 王吉耀. 循证医学与临床实践. 第2版. 北京:科技出版社, 2006
[47] Freeman AC, Sweeney K. Why general practitioners do not implement evidence: qualitative study. BMJ, 2001, 323: 1100
[48] Gabrielle van der Velde. Clinical decision analysis: an alternate, rigorous approach to making clinical decisions and developing treatment recommendations, 2005, 49(4): 258-263
[49] L McAuley, Ba' Pham, P Tugwell, D Moher. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet, 2000, 356(9237): 1228-1231
[50] Antomio Dans, Finlay Mcalister, Leonila Dans, et al. Therapy and applying the results. Applicability tools. INCLEN Knowledge Management Program, 2003
[51] Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clinical Chemistry, 2003, 49: 1-6
[52] Cohen H. How to write a patient case report. Am J Health Syst Pharm, 2006, 63(19): 1888-1892
[53] Davis P, Turner-Gomes S, Cunningham K, et al. Precision and accuracy of clinical and radiological signs in premature infants at risk of patent ductus arteriosus. Arch Pediatr Adolesc, 1995, 149(10): 1136-1141
[54] DiCenso A, Bayley L, Haynes B. Accessing pre-appraised evidence: fine tuning the 5S model into a 6S model. Evidence-based nursing, 2009, 12(4): 99-101
[55] Doi SA, Thalib L. A quality-effects model for meta-analysis. Epidemiology, 2008, 19(1): 94-100
[56] Guang Y, Xie L, Ding H, et al. Diagnosis value of focal liver lesions with SonoVue®-enhanced ultrasound compared with

contrast-enhanced computed tomography and contrast-enhanced MRI: a meta-analysis. J Cancer Res Clin Oncol, 2011, 137 (11): 1595-605

[57] Guyatt GH, Rennie D, Meade MO, Cook DJ. Users' Guides To The Medical Literature: A Manual for Evidence-Based Clinical Practice. AMA Press: Chicago, 2008

[58] Haynes RB. Of studies, syntheses, synopses, and systems: the "4S" evolution of services for finding current best evidence. ACP J Club, 2001, 134(2): A11-A13

[59] Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based health care decisions. ACP J Club, 2006, 145(3): A8

[60] Higgins JPT, Green S(editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

[61] Huwart L, Sempoux C, Vicaut E, et al. Magnetic resonance elastography for the noninvasive staging of liver fibrosis. Gastroenterology, 2008, 135(1): 32-40

[62] Jackson. R. Estimating cardiovascular risk and treatment benefit appendix I in Clinical Evidence, BMJ 2-3 ISSUE, 1999, 2000

[63] Jaeschke R, Guyatt G, Sckett D: User's guides to the medical literature Ⅲ. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA, 1994, 271(5): 389-391

[64] Jaeschke R, Guyatt G, Sckett DL: User's guides to the medical literature Ⅲ. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA, 1994, 271(9): 703-707

[65] Lau J, Antman EM, Jimenez-Silva J, et al. Cumulative Meta-Analysis of Therapeutic Trials for Myocardial Infarction. New England Journal of Medicine, 1992, 327(4): 248-254

[66] Li F, Wang Q, Zhang L, et al. The risk factors of transmission after the implementation of the routine immunization among children exposed to HBV infected mothers in a developing area in northwest China. Vaccine, 2012, 30(49): 7118-7122

[67] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions. Ann Intern Med, 2009, 151(4): W65-W94

[68] Ma Y, Wang X, Xu X, et al. World Health Organization subtypes, initial treatment outcome and prognostic study of unselected adult patients with acute myeloid leukaemia in Shanghai: an analysis of 623 cases. J Int Med Res, 2009, 37: 1191-1201

[69] McQueen M. Evidence-based laboratory medicine: addressing bias, generalisability and applicability in studies on diagnostic accuracy: the STARD initiative. Clin Biochem, 2003, 36: 1-2

[70] Miyawaki S, Ohtake S, Fujisawa S, et al. A randomized comparison of 4 courses of standard-dose multiagent chemotherapy versus 3 courses of high-dose cytarabine alone in postremission therapy for acute myeloid leukemia in adults: the JALSGAML201 Study. Blood, 2011, 117: 2366-2372

[71] Osterhuis WP, Bruns DE, Watine J et al. Evidence-Based Guidelines in Laboratory Medicine: Principles and Methods. Clinical Chemistry, 2004, 50: 806-818

[72] Price CP. Evidence-based laboratory medicine: Supporting decision-making, Clinical chemistry, 2000, 46: 1041-1050

[73] Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology. A basic science for clinical medicine. 2. The clinical examination. Little Boston: Brown and Company, 1985: 17-45

[74] Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. N Engl J Med, 1987, 316: 450-455

[75] Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials: an update of the quality and methodology. In: Bailar JC, Mosteller F, eds. Meidcal Uses of Statistics. 2nd ed. Boston, Mass: NEJM Books, 1992: 427-442

[76] Straus SE, Richardson WS, GlasziouP, Haynes RB. Evidence-Based Medicine: How to Practice and Teach it. Fourth Edition. Churchill Livingstone, 2011, 169-181

[77] Straus SE, Richardson WS, Paul Glasziou, Haynes RB. Evidence-based Medicine: How to Practice and Teach EBM, 3rd

Edition. Edinburgh: Churchill Livingstone, 2005

[78] Tsochatzis EA, Gurusamy KS, Ntaoula S, et al. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. J Hepatol, 2011, 54(4): 650-659

[79] Wang BQ, Zhu H, Liu HL, et al. Performance of Magnetic Resonance Elastography and Diffusion-Weighted Imaging for the Staging of Hepatic Fibrosis: A Meta-Analysis. Hepatology, 2012, 56: 239-247

[80] Wang XQ, Ryder J, Gross SA, et al. Prospective analysis of clinical and cytogenetic features of 435 cases of MDS diagnosed using the WHO(2001)classification: a prognostic scoring system for predicting survival in RCMD. Int J Hematol, 2009, 90: 361-369

[81] Wu B, Liu M, Zhang S. Dan Shen agents for acute ischaemic stroke. Cochrane Database of Systematic Reviews 2007, Issue 2. Art. No.: CD004295. DOI: 10.1002/14651858.CD004295.pub3

[82] Yanli Nie, Lin Li, Yurong Duan, Peixian Chen, Bruce H Barraclough, Mingming, Zhang and Jing Li. Patient safety education for undergraduate medical students: a systematic review. BMC Med Educ, 2011, 11: 33

[83] 王吉耀. 循证医学与临床实践. 第3版. 北京: 科学出版社, 2012

[84] 闫永平. 临床流行病学. 北京: 人民卫生出版社, 2009

[85] 林果为, 沈福民. 现代临床流行病学. 第2版. 上海: 复旦大学出版社, 2007

[86] 欧爱华, 老膺荣, 李晓庆, 等. 多中心临床随机对照试验的meta分析. 中华流行病学杂志, 2005, 26(4): 290-293

[87] 周立新, 谭家驹, 吴敏, 等. 广东省佛山市严重急性呼吸综合征首例报告. 中华结核病和呼吸杂志, 2010, 26(10): 598-601

[88] 黄悦勤. 预防医学. 北京: 北京医科大学出版社, 2004

[89] 谭燕, 刘鸣, 吴波. 葛根素治疗急性脑梗死的系统评价. 中国循证医学杂志, 2006, 6(7): 494-499

中英文名词对照索引

G

H

J

K

L

全国高等学校教材
供基础、临床、预防、口腔医学类专业用

全套教材均为卫生部“十二五”规划教材
全套教材（除 * 外）均为“十二五”普通高等教育本科国家级规划教材

策划编辑 左 巍 贾晓巍　封面设计 李 蹊 郭 淼
责任编辑 左 巍 孙雪冰　版式设计 赵京津 邹桂荣

本书附赠网络增值服务，激活方法：
1. 注册并登录人卫医学网教育频道（edu.ipmph.com）
2. 点击进入“网络增值服务”，搜索找到本书
3. 点击“激活”并输入“激活码”

人卫医学网 教育 edu.ipmph.com
刮开上网 享受增值

ISBN 978-7-117-17062-8
定 价：38.00 元